(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 November 2001 (01.11.2001)

PCT

(10) International Publication Number
WO 01/81588 A2

(51) International Patent Classification[7]: C12N 15/54, 9/12, 1/21, C07K 16/40, C12Q 1/68, G01N 33/68, A61K 38/43, 39/395, 31/7088

(21) International Application Number: PCT/US01/13784

(22) International Filing Date: 25 April 2001 (25.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/199,391 25 April 2000 (25.04.2000) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 01239 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **MEYERS, Rachel** [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

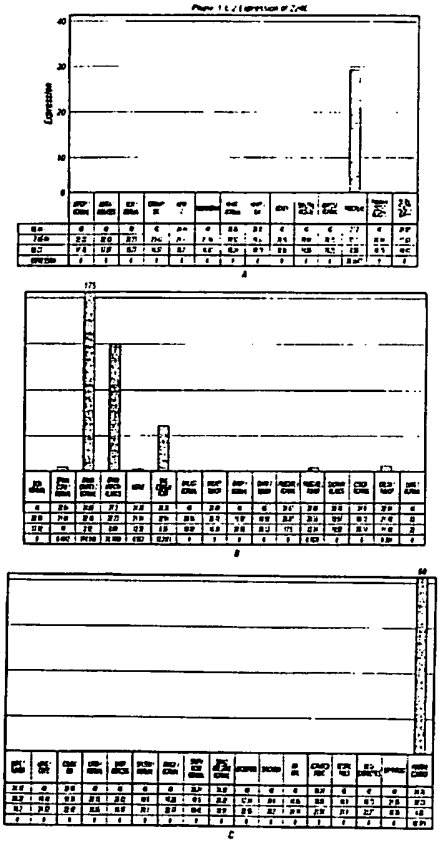
(74) Agents: **FAVORITO, Carolyn, A.** et al.; Morrison & Foerster LLP, Suite 500, 3811 Valley Centre Drive, San Diego, CA 92130-2332 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: 2246, NOVEL PROTEIN KINASE MOLECULES AND USES THEREFOR



(57) Abstract: The invention provides isolated nucleic acids molecules, designated 2246 nucleic acid molecules, which encode novel protein kinases. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 2246 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which 2246 gene has been introduced or disrupted. The invention still further provides isolated 2246 proteins, fusion proteins, antigenic peptides and anti-2246 antibodies. Diagnostic, screening and therapeutic methods utilizintg compositions of the invention are also provided.

WO 01/81588 A2

patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

2246, NOVEL PROTEIN KINASE MOLECULES AND USES THEREFOR

Background of the Invention

Phosphate tightly associated with protein has been known since the late nineteenth century. Since then, a variety of covalent linkages of phosphate to proteins have been found. The most common involve esterification of phosphate to serine, threonine, and tyrosine with smaller amounts being linked to lysine, arginine, histidine, aspartic acid, glutamic acid, and cysteine. The occurrence of phosphorylated proteins implies the existence of one or more protein kinases capable of phosphorylating amino acid residues on proteins, and also of protein phosphatases capable of hydrolyzing phosphorylated amino acid residues on proteins.

Kinases play a critical role in the mechanism of intracellular signal transduction. They act on the hydroxyamino acids of target proteins to catalyze the transfer of a high energy phosphate group from adenosine triphosphate (ATP). This process is known as protein phosphorylation. Along with phosphatases, which remove phosphates from phosphorylated proteins, kinases participate in reversible protein phosphorylation. Reversible phosphorylation acts as the main strategy for regulating protein activity in eukaryotic cells.

Protein kinases play critical roles in the regulation of biochemical and morphological changes associated with cell proliferation, differentiation, growth and division (D'Urso, G. *et al.* (1990) *Science* 250: 786-791; Birchmeier. C. *et al.* (1993) *Bioessays* 15: 185-189). They serve as growth factor receptors and signal transducers and have been implicated in cellular transformation and malignancy (Hunter, T. *et al.* (1992) *Cell* 70: 375-387; Posada, J. *et al.* (1992) *Mol. Biol. Cell* 3: 583-592; Hunter, T. *et al.* (1994) *Cell* 79: 573-582). For example, protein kinases have been shown to participate in the transmission of signals from growth-factor receptors (Sturgill, T. W. *et al.* (1988) *Nature* 344: 715-718; Gomez, N. *et al.* (1991) *Nature* 353: 170-173), control of entry of cells into mitosis (Nurse, P. (1990) *Nature* 344: 503-508; Maller, J. L. (1991) *Curr. Opin. Cell Biol.* 3: 269-275) and regulation of actin bundling (Husain-Chishti, A. *et al.* (1988) *Nature* 334: 718-721).

Kinases vary widely in their selectivity and specificity of target proteins. They still may, however, comprise the largest known enzyme superfamily. Protein kinases

can be divided into two main groups based on either amino acid sequence similarity or specificity for either serine/threonine or tyrosine residues. Serine/threonine specific kinases are often referred to as STKs while tyrosine specific kinases are referred to as PTKs. A small number of dual-specificity kinases are structurally like the serine/threonine-specific group. Within the broad classification, kinases can be further sub-divided into families whose members share a higher degree of catalytic domain amino acid sequence identity and also have similar biochemical properties. Most protein kinase family members also share structural features outside the kinase domain that reflect their particular cellular roles. These include regulatory domains that control kinase activity or interaction with other proteins (Hanks, S.K. *et al.* (1988) *Science* 241:42-52).

One example of the cellular role played by a STK is found in the case of *Halocynthia roretzi* posterior protein kinase-1 (HrPOPK-1). HrPOPK-1 contains a serine/threonine kinase domain and apparently plays a role in embryogenesis, likely in part via temporal regulation of its expression. HrPOPK-1 mRNA has been localized at the posterior pole of embryos and shows zygotic expression in neural tissues at the tailbud stage. (Sasakura, Y. *et al.* (1998) "Maternally localized RNA encoding a serine/threonine protein kinase in the ascidian, *Halocynthia roretzi*," *Mechanisms of Development* 76:161-163.)

Almost all kinases contain a catalytic domain composed of 250-300 conserved amino acids. This catalytic domain may be viewed as composed of 11 subdomains. Some of these subdomains apparently contain distinct amino acid motifs which confer specificity as a STK or PTK or both. Kinases may also contain additional amino acid sequences, usually between 5 and 100 residues, flanking or occurring within the catalytic domain. These residues apparently act to regulate kinase activity and to determine substrate specificity. (Reviewed in Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I:7-20 Academic Press, San Diego, Calif.)

Approximately one third of the known oncogenes encode PTKs. PTKs may occur as either transmembrane or soluble proteins. Transmembrane PTKs act as receptors for many growth factors. Interaction of a growth factor to its cognate receptor initiates the phosphorylation of specific tyrosine residues in the receptor itself as well as in certain second messenger proteins. Growth factors found to associate

with such PTK receptors include epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, hepatocyte growth factor, insulin and insulin-like growth factors, nerve growth factor, vascular endothelial growth factor, and macrophage colony stimulating factor.

Soluble PTKs often interact with the cytosolic domains of plasma membrane receptors. Receptors that signal through such PTKs include cytokine, hormone, and antigen-specific lymphocytic receptors. Many PTKs were identified as oncogene products by the observation that PTK activation was no longer subject to normal cellular controls. Also, increased tyrosine phosphorylation activity is often observed in cellular transformation, or oncogenesis, (Carbonneau, H. and Tonks, N. K. (1992) Annu. Rev. Cell Biol. 8:463-93.) PTK regulation may therefore be an important strategy in controlling some types of cancer.

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel nucleic acid molecules and proteins encoded by such nucleic acid molecules, referred to herein as "kinases" or by the individual clone names "2246". The 2246 nucleic acid and protein molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, e.g., including cell proliferation, differentiation, growth and division. In particular, the kinase and its related nucleic acids will be advantageous in the regulation of any cellular function uncontrolled proliferation and differentiation, such as in cases of cancer. Other situations where the kinases of the invention are of particular advantage are in cases of autoimmune disorders or undesired inflammation. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding 2246 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of 2246-encoding nucleic acids.

In one embodiment, a 2246 nucleic acid molecule is 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 90%, 95%, 98% homologous to a nucleotide sequence including SEQ ID NO:1, SEQ ID NO:3, or a complement thereof.

In another embodiment, a 2246 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous

to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment, a 2246 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to an amino acid sequence including SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of a human 2246. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein which includes the amino acid sequence of SEQ ID NO:2. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO:2.

Another embodiment of the invention features nucleic acid molecules, preferably 2246 nucleic acid molecules, which specifically detect 2246 nucleic acid molecules relative to nucleic acid molecules encoding non-2246 proteins. For example, in one embodiment, such a nucleic acid molecule is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 800 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, or a complement thereof.

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide which includes the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule which includes SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a 2246 nucleic acid molecule, e.g., the coding strand of a 2246 nucleic acid molecule.

Another aspect of the invention provides a vector comprising a 2246 nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. The invention also provides a method for producing a protein, preferably a 2246 protein, by culturing in a suitable medium, a host cell, e.g., a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

Another aspect of this invention features isolated or recombinant 2246 proteins and polypeptides.

In one embodiment, the isolated protein, preferably a 2246 protein, includes at least one Ser/Thr kinase site. In another embodiment, the isolated protein, preferably a 2246 protein, includes at least one Ser/Thr kinase site and has an amino acid sequence which is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to an amino acid sequence including SEQ ID NO:2. In an even further embodiment, the isolated protein, preferably a 2246 protein, includes at least one Ser/Thr kinase site and plays a role in signaling pathways associated with cellular growth, e.g., signaling pathways associated with cell cycle regulation. In another embodiment, the isolated protein, preferably a 2246 protein, includes at least one Ser/Thr kinase site and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3.

In another embodiment, the isolated protein, preferably a 2246 protein, has an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment, the protein, preferably a 2246 protein, has an amino acid sequence at least about 50%, 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or more homologous to an amino acid sequence including SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2). In another embodiment, the invention features fragments of the proteins having the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 amino acids (e.g., contiguous amino acids) of the amino acid sequence of SEQ ID NO:2, respectively. In another embodiment, the protein, preferably a 2246 protein, has the amino acid sequence of SEQ ID NO:2.

Another embodiment of the invention features an isolated protein, preferably a 2246 protein, which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 50%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 90%, 95%, 97%, 98% or more homologous to a nucleotide sequence (e.g., to the entire length of the nucleotide sequence) including SEQ ID NO:1, SEQ ID NO:3, or a complement thereof. This invention further features an isolated protein, preferably a 2246 protein, which is encoded by a nucleic acid molecule having a

nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or a complement thereof.

The proteins of the present invention or biologically active portions thereof, can be operatively linked to a non-2246 polypeptide (e.g., heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably 2246 proteins. In addition, the 2246 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of a 2246 nucleic acid molecule, protein or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting a 2246 nucleic acid molecule, protein or polypeptide such that the presence of a 2246 nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of 2246 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of 2246 activity such that the presence of 2246 activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating 2246 activity comprising contacting a cell capable of expressing 2246 with an agent that modulates 2246 activity such that 2246 activity in the cell is modulated. In one embodiment, the agent inhibits 2246 activity. In another embodiment, the agent stimulates 2246 activity. In one embodiment, the agent is an antibody that specifically binds to a 2246 protein. In another embodiment, the agent modulates expression of 2246 by modulating transcription of a 2246 gene or translation of a 2246 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a 2246 mRNA or a 2246 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant 2246 protein or nucleic acid expression or activity by administering an agent which is a 2246 modulator to the subject. In one embodiment, the 2246 modulator is a 2246 protein. In another

embodiment the 2246 modulator is a 2246 nucleic acid molecule. In yet another embodiment, the 2246 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant 2246 protein or nucleic acid expression is a cellular growth related disorder.

Disorders associated with the following cells or tissues, in which 2246 has been expressed as shown in Figure 6, are also encompassed: pancreas, brain cortex, brain hypothalamus, nerve, and DRG, among others. Other disorders are those characterized by aberrant activity or expression of the 2246 polypeptides or nucleic acids, as well as aberrant or deficient mobilization of an intracellular molecule that participates in a phosphorylation; and/or aberrant or deficient modulation of function, survival, morphology, proliferation and/or differentiation of cells of tissues in which 2246 molecules are expressed.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a 2246 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a 2246 protein, wherein a wild-type form of the gene encodes a protein with a 2246 activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a 2246 protein, by providing an indicator composition comprising a 2246 protein having 2246 activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on 2246 activity in the indicator composition to identify a compound that modulates the activity of a 2246 protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1a-c depict a cDNA sequence (SEQ ID NO:1) and predicted amino acid sequence (SEQ ID NO:2) of human 2246. The location of the methionine-initiated open reading frame of human 2246 (without the 5' and 3' untranslated regions) is also indicated in the Figure (SEQ ID NO:3).

Figure 2 depicts a hydropathy plot of human 2246. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The location of the transmembrane domains and the extracellular and intracellular loops is also indicated. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 2246 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, e.g., a sequence above the dashed line, e.g., the sequence from about amino acid 205 to 215, from about 545 to 560, and from about 610 to 620 of SEQ ID NO:2; all or part of a hydrophilic sequence, e.g., a sequence below the dashed line, e.g., the sequence from about amino acid 340 to 380, from about 430 to 470, and from about 500 to 520 of SEQ ID NO:2; a sequence which includes a Cys, or a glycosylation site.

Figure 3 depicts an alignment of the eukaryotic protein kinase family domain of human 2246 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequence is the consensus amino acid sequence (SEQ ID NO:4), while the lower amino acid sequence corresponds to amino acids 19 to 270 of SEQ ID NO:2.

Figures 4a-b depict a BLAST alignment of human 2246 with a consensus amino acid sequence derived from a ProDomain "HRPOPK-1" (Release 1999.2; see also ProDomain Release 2000.1; http://www.toulouse.inra.fr/prodom.html). The lower sequence is amino acid residues 1 to 227 of the amino acid consensus sequence (SEQ ID NOs:5 and 6), while the upper amino acid sequence corresponds to the "HRPOPK-1" domain of human 2246, amino acid residues 282 to 373 and 475 to 507 of SEQ ID NO:2. The BLAST algorithm identifies multiple local alignments between the consensus amino acid sequence and human 2246. Figure 4A depicts the first local alignment and Figure 4B the second.

Figure 5 depicts a BLAST alignment of human 2246 with a consensus amino acid sequence derived from a ProDomain "HRPOPK-1" (Release 1999.2; see also ProDomain Release 2000.1; http://www.toulouse.inra.fr/prodom.html). The lower sequence is amino acid residues 1 to 125 of the 125 amino acid consensus sequence

(SEQ ID NO:7), while the upper amino acid sequence corresponds to the "HRPOPK-1" domain of human 2246, amino acid residues 521 to 627 of SEQ ID NO:2.

Figure 6 depicts a BLAST alignment of human 2246 with a consensus amino acid sequence derived from a ProDomain "ATP-binding serine/threonine protein - phosphorylation receptor tyrosine - protein precursor transmembrane" (Release 1999.2; see also ProDomain Release 2000.1; http://www.toulouse.inra.fr/prodom.html). The lower sequence is amino acid residues 183 to 249 of the 67 amino acid consensus sequence (SEQ ID NOs:8-11), while the upper amino acid sequence corresponds to the "ATP-binding serine/threonine protein - phosphorylation receptor tyrosine - protein precursor transmembrane" domain of human 2246, amino acid residues 113 to 163, 178 to 216, 21 to 98, and 239 to 271 of SEQ ID NO:2. The BLAST algorithm identifies multiple local alignments between the consensus amino acid sequence and human 2246. Figure 6A depicts the first local alignment, Figure 6B the second, Figure 6C the third, and Figure 6D the fourth.

Figure 7 is a panel bar graph depicting the relative expression of 2246 RNA relative to a no template control in a panel of human tissues or cells, including but not limited to normal artery, diseased aorta, normal vein, coronary smooth muscle cells (SMC) human umbilical vein endothelial cells (HUVEC), hemangioma, normal heart, coronary heart failure heart tissue, skeletal muscle, kidney, normal adipose, pancreas, primary osteoblasts, osteoclasts, skin, spinal cord, brain cortex, brain hypothalamus, nerve, dorsal root ganglia (DRS), normal breast, breast tumor, normal ovary, ovary tumor, normal prostate and prostate tumor, salivary glands, normal colon and colon tumor, normal lung and lung tumor, lung COPD, colon IBD, normal liver and liver fibrosis, spleen, tonsil, lymph node, small intestine, macrophages, synovium, BM-MNC, activated PBMC, neutrophlis, megakaryocytes, and erythroid, among others, detected using real-time quantitative RT-PCR Taq Man analysis. The graph indicates significant 2246 expression in normal human brain cortex, as well as high expression in pancreas and brain hypothalamus.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as "2246" nucleic acid and polypeptide molecules, which

play a role in or function in signaling pathways associated with cellular growth. In one embodiment, the 2246 molecules modulate the activity of one or more proteins involved in cellular growth or differentiation, e.g., cardiac cell growth or differentiation. In another embodiment, the 2246 molecules of the present invention are capable of modulating the phosphorylation state of a 2246 molecule or one or more proteins involved in cellular growth or differentiation.

As used herein, the term "protein kinase" includes a protein or polypeptide which is capable of modulating its own phosphorylation state or the phosphorylation state of another protein or polypeptide. Protein kinases can have a specificity for (i.e., a specificity to phosphorylate) serine/threonine residues, tyrosine residues, or both serine/threonine and tyrosine residues, e.g., the dual specificity kinases. As referred to herein, protein kinases preferably include a catalytic domain of about 200-400 amino acid residues in length, preferably about 200-300 amino acid residues in length, or more preferably about 250-300 amino acid residues in length, which includes preferably 5-20, more preferably 5-15, or preferably 11 highly conserved motifs or subdomains separated by sequences of amino acids with reduced or minimal conservation. Specificity of a protein kinase for phosphorylation of either tyrosine or serine/threonine can be predicted by the sequence of two of the subdomains (VIb and VIII) in which different residues are conserved in each class (as described in, for example, Hanks *et al.* (1988) *Science* 241:42-52) the contents of which are incorporated herein by reference). These subdomains are also described in further detail herein.

Protein kinases play a role in signaling pathways associated with cellular growth. For example, protein kinases are involved in the regulation of signal transmission from cellular receptors, e.g., growth-factor receptors; entry of cells into mitosis; and the regulation of cytoskeleton function, e.g., actin bundling. Thus, the 2246 molecules of the present invention may be involved in: 1) the regulation of transmission of signals from cellular receptors, e.g., cardiac cell growth factor receptors; 2) the modulation of the entry of cells into mitosis; 3) the modulation of cellular differentiation; 4) the modulation of cell death; and 5) the regulation of cytoskeleton function, e.g., actin bundling.

Inhibition or over stimulation of the activity of protein kinases involved in signaling pathways associated with cellular growth can lead to perturbed cellular growth, which can in turn lead to cellular growth related disorders. As used herein, a "cellular growth related disorder" includes a disorder, disease, or condition characterized by a deregulation, e.g., an upregulation or a downregulation, of cellular growth. Cellular growth deregulation may be due to a deregulation of cellular proliferation, cell cycle progression, cellular differentiation and/or cellular hypertrophy.

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as 2246 protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin. Members of a family may also have common functional characteristics.

One embodiment of the invention features 2246 nucleic acid molecules, preferably human 2246 molecules, e.g., 2246. The 2246 nucleic acid and protein molecules of the invention are described in further detail in the following subsections.

A. The 2246 Nucleic Acid and Protein Molecules

The human 2246 sequence (Figure 1; SEQ ID NO:1), which is approximately 2219 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2025 nucleotides, including the termination codon (nucleotides indicated as the coding region of SEQ ID NO:1 in Fig. 1; SEQ ID NO:3). The coding sequence encodes a 674 amino acid protein (SEQ ID NO:2).

Human 2246 contains the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997) Protein 28:405-420 and http://www.psc.edu/general/software/packages/pfam/pfam.html):

a eukaryotic protein kinase domain (PFAM Accession Number PF00069) located at about amino acid residues 19 to 270 of SEQ ID NO:2;

two dileucine motifs in the tail (LL) (predicted by PSORT), Nakai, K. and Kanehisa, M. (1992) *Genomics* 14:897-911) at about amino acids 147 to 148 and 148 to 149;

two transmembrane domains (predicted by MEMSAT, Jones et al. (1994) Biochemistry 33:3038-3049) at about amino acids 199 to 216 and 550 to 566 of SEQ ID NO:2;

one cAMP- and cGMP-dependent protein kinase phosphorylation site (Prosite PS00004) from about amino acids 257 to 260 of SEQ ID NO:2;

eleven protein kinase C phosphorylation sites (Prosite PS00005) at about amino acids 4 to 6, 114 to 116, 427 to 429, 435 to 437, 443 to 445, 451 to 453, 481 to 483, 489 to 491, 545 to 547, 569 to 571 and 619 to 621 of SEQ ID NO:2;

sixteen casein kinase II phosphorylation sites (Prosite PS00006) located at about amino acids 4 to 7, 99 to 102, 114 to 117, 127 to 130, 170 to 173, 251 to 254, 294 to 297, 324 to 327, 346 to 349, 387 to 390, 451 to 454, 501 to 504, 512 to 515, 594 to 597, 635 to 638, and 668 to 671, of SEQ ID NO:2;

one tyrosine kinase phosphorylation site (Prosite PS00007) located at about amino acids 588 to 595;

five N-myristoylation sites (Prosite PS00008) from about amino acids 9 to 14, 415 to 420, 523 to 528, 654 to 659, and 663 to 668 of SEQ ID NO:2;

two amidation sites (Prosite PS00009) from about amino acids 193 to 196 and 373 to 376 of SEQ ID NO:2; and

one serine/threonine protein kinase active-site signal (Prosite PS00108) from about amino acids 137 to 149 of SEQ ID NO:2.

As used herein, the term "kinase domain" includes an amino acid sequence of about 100 to 275 amino acid residues in length and having a bit score for the alignment of the sequence to the kinase domain (HMM) of at least 100. Preferably a

kinase domain mediates intracellular signal transduction. Preferably, a kinase domain includes at least about 100 to 275 amino acids, more preferably about 150 to 275 amino acid residues, or about 200 to 275 amino acids and has a bit score for the alignment of the sequence to the kinase domain (HMM) of at least 100, 150, 200, 250 or greater. An alignment of the kinase domain (amino acids 19 to 270 of SEQ ID NO:2) of human 2246 with a consensus amino acid sequence (SEQ ID NO:2) derived from a hidden Markov model is depicted in Figure 3. The "protein kinase" domain (HMM) has been assigned the PFAM Accession Number PF00069 (http://genome.wustl.edu/Pfam/.html) and corresponds to about amino acids 19 to 270 of SEQ ID NO:2.

In a preferred embodiment, a 2246 polypeptide or protein has a "kinase domain" or a region which includes at least about 100 to 215 more preferably about 150 to 275 or 200 to 275 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "kinase domain," e.g., the kinase domain of human 2246 (e.g., residues 19 to 270 of SEQ ID NO:2).

To identify the presence of a "kinase" domain in a 2246 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the Pfam database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer et al. (1997) Proteins 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov et al. (1990) Meth. Enzymol.183:146-159; Gribskov et al. (1987) Proc. Natl. Acad. Sci. USA 84:4355-4358; Krogh et al. (1994) J. Mol. Biol. 235:1501-1531; and Stultz et al. (1993) Protein Sci. 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a "kinase domain" domain in the amino acid sequence of human 2246 at about residues 19 to 270 of SEQ ID NO:2 (see Figure 1).

To identify the presence of a "kinase" domain in a 2246 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of domains, e.g., the ProDom database (Corpet et al. (1999), Nucl. Acids Res. 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-BLAST searches (Altschul SF et al. (1997) Nucleic Acids Res. 25:3389-3402; Gouzy et al. (1999) Computers and Chemistry 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain. A BLAST search was performed against the HMM database resulting in the identification of a "kinase" domain in the amino acid sequence of human 2246 at about residues 282 to 373, 475 to 507, 521 to 627 and 21 to 271 of SEQ ID NO:2 (see Figure 1).

The kinase domain is homologous to ProDom family "HRPOPK-1," SEQ ID NOs:5 and 6, (ProDomain Release 1999.2 http://www.toulouse.inra.fr/prodom.html). The consensus sequence for SEQ ID NO:5 is 43% identical over amino acids 282 to 373 of SEQ ID NO:2 as shown in Figure 4a and the consensus sequence for SEQ ID NO:6 is 42% identical over amino acids 475 to 507 of SEQ ID NO:2 as shown in Figure 4b.

The kinase domain is also homologous to ProDom family "HRPOPK-1," SEQ ID NO:7, (ProDomain Release 1999.2 http://www.toulouse.inra.fr/prodom.html). The consensus sequence for SEQ ID NO:7 is 37% identical over amino acids 521 to 627 of SEQ ID NO:2 as shown in Figure 5.

The kinase domain is homologous to ProDom family "ATP-binding serine/threonine protein - phosphorylation receptor tyrosine - protein precursor transmembrane," SEQ ID NOs:8 to 11, (ProDomain Release 1999.2 http://www.toulouse.inra.fr/prodom.html). The consensus sequence for SEQ ID NO:8 is 41% identical over amino acids 113 to 163 of SEQ ID NO:2 as shown in Figure 6a; the consensus sequence for SEQ ID NO:9 is 46% identical over amino acids 178 to 216 of SEQ ID NO:2 as shown in Figure 6b; the consensus sequence for SEQ ID NO:10 is 31% identical over amino acids 21 to 98 of SEQ ID NO:2 as shown in

Figure 6c; the consensus sequence for SEQ ID NO:11 is 29% identical over amino acids 239 to 271 of SEQ ID NO:2 as shown in Figure 6d.

In another embodiment, the isolated proteins of the present invention, preferably 2246 proteins, are identified based on the presence of at least one Ser/Thr kinase site.

As used herein, the term "Ser/Thr kinase site" includes an amino acid sequence of about 200-400 amino acid residues in length, preferably 200-300 amino acid residues in length, and more preferably 250-300 amino acid residues in length, which is conserved in kinases which phosphorylate serine and threonine residues and found in the catalytic domain of Ser/Thr kinases. Preferably, the Ser/Thr kinase site includes the following amino acid consensus sequence X_9-g-X-G-X_4-V-X_{12}-K-X-(10-19)-E-X_{66}-h-X_8-h-r-D-X-K-X_2-N-X_{17}-K-X_2-D-f-g-X_{21}-p-X_{13}-w-X_3-g-X_{55}-R-X_{14}-h-X_3 (SEQ ID NO:12) (where invariant residues are indicated by upper case letters and nearly invariant residues are indicated by lower case letters). The nearly invariant residues are usually found in most Ser/Thr kinase sites, but can be replaced by other amino acids which, preferably, have similar characteristics. For example, a nearly invariant hydrophobic amino acid in the above amino acid consensus sequence would most likely be replaced by another hydrophobic amino acid. Ser/Thr kinase domains are described in, for example, Levin D.E. *et al.* (1990) *Proc. Natl. Acad. Sci.* USA 87:8272-76, the contents of which are incorporated herein by reference.

In a preferred embodiment, the 2246 includes the following Prosite signature (PS00108) amino acid consensus sequence for a serine/threonine protein kinase active-site signal, or sequence homologous thereto: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT] (SEQ. ID. NO:13). In the above conserved motif, and other motifs described herein, the standard IUPAC one-letter code for the amino acids is used. Each element in the pattern is separated by a dash (-); square brackets ([ ]) indicate the particular residues that are accepted at that position; x indicates that any residue is accepted at that position; and numbers in parentheses (()) indicate the number of residues represented by the accompanying amino acid.

A 2246 polypeptide can include at least one "transmembrane domain" or region homologous with a "transmembrane domain". As used herein, the term "transmembrane domain" includes an amino acid sequence of about 10 to 40 amino acid residues in

length and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, e.g., at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains typically have alpha-helical structures and are described in, for example, Zagotta, W.N. et al., (1996) *Annual Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference.

In a preferred embodiment, a 2246 polypeptide or protein has at least one, preferably two "transmembrane domains" or regions which include at least about 10 to 35 more preferably about 10 to 25 or 15 to 20 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "transmembrane domain," e.g., the transmembrane domains of human 2246 (e.g., residues 199 to 216 and 550 to 566 of SEQ ID NO:2). The transmembrane domain of human 2246 is visualized in the hydropathy plot (Figure 2) as regions of about 15 to 20 amino acids where the hydropathy trace is mostly above the horizontal line.

To identify the presence of a "transmembrane" domain in a 2246 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be analyzed by a transmembrane prediction method that predicts the secondary structure and topology of integral membrane proteins based on the recognition of topological models (MEMSAT, Jones et al., (1994) *Biochemistry* 33:3038-3049).

A 2246 polypeptide can include at least one, two, preferably three "non-transmembrane regions." As used herein, the term "non-transmembrane region" includes an amino acid sequence not identified as a transmembrane domain. The non-transmembrane regions in 2246 are located at about amino acids 1 to 198, 217 to 549 and 567 to 674 of SEQ ID NO:2.

The non-transmembrane regions of 2246 include at least one cytoplasmic region. When located at the N-terminus, the cytoplasmic region is referred to herein as the "N-terminal cytoplasmic domain." As used herein, an "N-terminal cytoplasmic domain" includes an amino acid sequence having about 1 to 250, preferably about 1 to 225, more preferably about 1 to 200, or even more preferably about 1 to 198 amino acid residues in length and is located inside of a cell or within the cytoplasm of a cell. The C-terminal amino acid residue of an "N-terminal cytoplasmic domain" is adjacent to a N-terminal

amino acid residue of a transmembrane domain in a 2246 protein. For example, N-terminal cytoplasmic domains are located at about amino acid residues 1 to 198 of SEQ ID NO:2.

In a preferred embodiment, a polypeptide or protein has an N-terminal cytoplasmic domain or a region which includes at least about 50, preferably about 1 to 75, and more preferably about 1 to 198 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "N-terminal cytoplasmic domain," e.g., the N-terminal cytoplasmic domain of human 2246 (e.g., residues 1 to 198 of SEQ ID NO:2).

In another embodiment, a cytoplasmic region of a 2246 protein can include the C-terminus and can be a "C-terminal cytoplasmic domain," also referred to herein as a "C-terminal cytoplasmic tail." As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 50, preferably about 1 to 100, more preferably about 1 to 108 amino acid residues and is located inside of a cell or within the cytoplasm of a cell. The N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a 2246 protein. For example, a C-terminal cytoplasmic domain is located at about amino acid residues 567 to 674 of SEQ ID NO:2.

In a preferred embodiment, a 2246 polypeptide or protein has a C-terminal cytoplasmic domain or a region which includes at least about 5, preferably about 5 to 100, and more preferably about 1 to 108 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a C-terminal cytoplasmic domain," e.g., the C-terminal cytoplasmic domain of human 2246 (e.g., residues 567 to 674 of SEQ ID NO:2).

In another embodiment, a 2246 protein includes at least one non-cytoplasmic loop. As used herein, a "non-cytoplasmic loop" includes an amino acid sequence located outside of a cell or within an intracellular organelle. Non-cytoplasmic loops include extracellular domains (i.e., outside of the cell) and intracellular domains (i.e., within the cell). When referring to membrane-bound proteins found in intracellular organelles (e.g., mitochondria, endoplasmic reticulum, peroxisomes microsomes, vesicles, endosomes, and lysosomes), non-cytoplasmic loops include those domains of the protein that reside in the lumen of the organelle or the matrix or the intermembrane space. For example, a

"non-cytoplasmic loop" can be found at about amino acid residues 217 to 549 of SEQ ID NO:2.

In a preferred embodiment, a 2246 polypeptide or protein has at least one non-cytoplasmic loop or a region which includes at least about 4, preferably at least about 5 to 200, more preferably at least about 6 to 300 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "non-cytoplasmic loop," e.g., at least one non-cytoplasmic loop of human 2246 (e.g., residues list all 217 to 549 of SEQ ID NO:2).

A 2246 family member can include at least one protein kinase domain, at least one, two or preferably three transmembrane or non-transmembrane domains, or at least one, preferably two dileucine motifs. Furthermore, a 2246 family member can include at least one cAMP- and cGMP-dependent protein kinase phosphorylation site (Prosite PS00004); at least one, two, three, four, five, six, seven, eight, nine, ten and preferably eleven protein kinase C phosphorylation sites (PS00005); at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, and preferably sixteen casein kinase II phosphorylation sites (Prosite PS00006); at least one tyrosine kinase phosphorylation site (Prosite PS00007); at least one, two, three, four, and preferably five N-myristoylation sites (PS00008); at least one protein kinases ATP-binding region signature (Prosite PS00107); at least one and preferably two amidation sites (Prosite PS00009); or at least one serine/threonine protein kinase active-site signal site (Prosite PS00108).

Isolated proteins of the present invention, preferably 2246 proteins, have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2 or are encoded by a nucleotide sequence sufficiently homologous to SEQ ID NO:1 or SEQ ID NO:3. The 2246 nucleic acid encodes a polypeptide with homology to a previously identified HrPOPK-1 kinases having a Ser/Thr domain. Thus the 2246 encoded polypeptide is expected to be a kinase and function in the phosphorylation of protein substrates. Additionally, the 2246 kinase is expected to have additional biological activities similar to HrPOPK-1 kinase described above. These include possible roles in regulating embryogenesis, possibly via appropriate localization of 2246 encoding mRNA in embryos, and/or regulating development of neural tissues in zygotes. Additionally, 2246 encoded proteins may be used in cases of deficiencies in HrPOPK-1

like activity, such as those resulting in defective embryogenesis or development of neural tissues.

As used interchangeably herein a "2246 activity", "biological activity of 2246" or "functional activity of 2246", refers to an activity exerted by a 2246 protein, polypeptide or nucleic acid molecule on a 2246 responsive cell or a 2246 protein substrate as determined *in vivo*, or *in vitro*, according to standard techniques. The biological activity of 2246 is described herein.

Thus, the 2246 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more disorders. Examples of such disorders, e.g., kinase-associated or other 2246-associated disorders, include but are not limited to, cellular proliferative and/or differentiative disorders, disorders associated with bone metabolism, immune e.g., inflammatory, disorders, cardiovascular disorders, including endothelial cell disorders, liver disorders, viral diseases, pain or metabolic disorders.

Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of ovary, prostate, colon, lung, breast and liver origin.

As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, e.g., malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, e.g., cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell

carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

The 2246 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L. (1991) Crit Rev. in Oncol./Hemotol. 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

Aberrant expression and/or activity of 2246 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect

effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 2246 molecules effects in bone cells, e.g. osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, 2246 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 2246 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyrodism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

The 2246 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, e.g., inflammatory, (e.g. respiratory inflammatory) disorders. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosis, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease, e.g. Crohn's disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves'

disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

Examples of disorders involving the heart or "cardiovascular disorder" include, but are not limited to, a disease, disorder, or state involving the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, e.g., by a thrombus. Examples of cardiovascular disorders include but are not limited to, hypertension, atherosclerosis, coronary artery spasm, coronary artery disease, arrhythmias, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus,

tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, disorders involving cardiac transplantation, and congestive heart failure.

A cardiovascular disease or disorder also includes an endothelial cell disorder.

As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, e.g., proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (e.g., atherosclerosis), and chronic inflammatory diseases (e.g., rheumatoid arthritis).

Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein can be used for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate,

isonizaid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

Additionally, 2246 molecules can play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 2246 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 2246 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

Additionally, 2246 can play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L. (1987) Pain, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

Accordingly, another embodiment of the invention features isolated 2246 proteins and polypeptides having a 2246 activity. Preferred proteins are 2246 proteins having at least one Ser/Thr kinase site. Additional preferred proteins have at least one Ser/Thr kinase site and preferably a 2246 activity. Additional preferred proteins have at least one Ser/Thr kinase site and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3.

The nucleotide sequence of the isolated human 2246 cDNA and the predicted amino acid sequence of the human 2246 polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding human 2246 was deposited with American Type Culture Collection

(ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human 2246 gene, which is approximately 2219 nucleotides in length, encodes a protein having a molecular weight of approximately 36 kD and which is approximately 674 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode 2246 proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify 2246-encoding nucleic acids (e.g., 2246 mRNA) and fragments for use as PCR primers for the amplification or mutation of 2246 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated or purified" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated 2246 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in

genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence of SEQ ID NO:1, or the nucleotide sequence of SEQ ID NO:3, as a hybridization probe, nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1 or SEQ ID NO:3 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1 or SEQ ID NO:3, respectively.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to 2246 nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the partial human 2246 cDNA. This cDNA comprises sequences encoding the partial human 2246 protein (i.e., "the coding region", as shown in SEQ ID NO:3), as well as 5' untranslated sequences (52 nucleotides before the coding region) and 3' untranslated sequences (142 nucleotides after the coding

region). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:3 (e.g., corresponding to SEQ ID NO:3).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, respectively, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, respectively, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 90%, 95%, 97%, 98% or more homologous to the nucleotide sequence (e.g., to the entire length of the nucleotide sequence) shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a 2246 protein. The nucleotide sequence determined from the cloning of the 2246 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 2246 family members, as well as 2246 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1 or SEQ ID NO:3, of an anti-sense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3. In an exemplary embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least 350, 400, 450, 500, 550, 600, 650, 700, 750, or

800 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1 or SEQ ID NO:3.

Probes based on the 2246 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which misexpress a 2246 protein, such as by measuring a level of a 2246-encoding nucleic acid in a sample of cells from a subject e.g., detecting 2246 mRNA levels or determining whether a genomic 2246 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a 2246 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, which encodes a polypeptide having a 2246 biological activity (the biological activities of the 2246 proteins are described herein), expressing the encoded portion of the 2246 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 2246 protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, due to the degeneracy of the genetic code and, thus, encode the same 2246 proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2.

In addition to the 2246 nucleotide sequences shown in SEQ ID NO:1 or SEQ ID NO:3, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the 2246 proteins may exist within a population (e.g., the human population). Such genetic polymorphism in the 2246 genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding an 2246 protein, preferably a mammalian 2246 protein, and can further include non-coding regulatory sequences, and introns. Such natural allelic variations include both

functional and non-functional 2246 proteins and can typically result in 1-5% variance in the nucleotide sequence of a 2246 gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in 2246 genes that are the result of natural allelic variation and that do not alter the functional activity of a 2246 protein are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding other 2246 family members and, thus, which have a nucleotide sequence which differs from the 2246 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention. For example, another 2246 cDNA can be identified based on the nucleotide sequence of human 2246. Moreover, nucleic acid molecules encoding 2246 proteins from different species, and thus which have a nucleotide sequence which differs from the 2246 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention. For example, a mouse 2246 cDNA can be identified based on the nucleotide sequence of a human 2246.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the 2246 cDNAs of the invention can be isolated based on their homology to the 2246 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3. In other embodiment, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-

6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or SEQ ID NO:3 corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the 2246 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1 or SEQ ID NO:3, thereby leading to changes in the amino acid sequence of the encoded 2246 proteins, without altering the functional ability of the 2246 proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1 or SEQ ID NO:3. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 2246 (e.g., the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the 2246 proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 2246 proteins of the present invention and other 2246 family members are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding 2246 proteins that contain changes in amino acid residues that are not essential for activity. Such 2246 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 41%, 42%, 45%, 50%, 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or more homologous to the

amino acid sequence of SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

An isolated nucleic acid molecule encoding a 2246 protein homologous to the protein of SEQ ID NO:2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, respectively, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 2246 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 2246 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 2246 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant 2246 protein can be assayed for the ability to: 1) regulate transmission of signals from cellular receptors, e.g., cardiac cell growth factor receptors; 2) control entry of cells into mitosis; 3) modulate cellular differentiation; 4) modulate cell death; or 5) regulate cytoskeleton function, e.g., actin bundling.

In addition to the nucleic acid molecules encoding 2246 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire 2246 coding strand, or only to a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding 2246. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of human 2246 corresponds to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 2246. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding 2246 disclosed herein (e.g., SEQ ID NO:3), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of 2246 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 2246 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 2246 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be

used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 2246 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic

acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave 2246 mRNA transcripts to thereby inhibit translation of 2246 mRNA. A ribozyme having specificity for a 2246-encoding nucleic acid can be designed based upon the nucleotide sequence of a 2246 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 2246-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 2246 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, 2246 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 2246 (e.g., the 2246 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 2246 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug*

Des. 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the 2246 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* Proc. Natl. Acad. Sci. 93: 14670-675.

PNAs of 2246 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 2246 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of 2246 can be modified, (e.g., to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of 2246 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (e.g., RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked

using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

II. Isolated 2246 Proteins and Anti-2246 Antibodies

One aspect of the invention pertains to isolated 2246 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-2246 antibodies. In one embodiment, native 2246 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, 2246 proteins are produced by recombinant DNA techniques. Alternative to recombinant

expression, a 2246 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the 2246 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of 2246 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of 2246 protein having less than about 30% (by dry weight) of non-2246 protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-2246 protein, still more preferably less than about 10% of non-2246 protein, and most preferably less than about 5% non-2246 protein. When the 2246 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of 2246 protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of 2246 protein having less than about 30% (by dry weight) of chemical precursors or non-2246 chemicals, more preferably less than about 20% chemical precursors or non-2246 chemicals, still more preferably less than about 10% chemical precursors or non-2246 chemicals, and most preferably less than about 5% chemical precursors or non-2246 chemicals.

Biologically active portions of a 2246 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 2246 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 2246 proteins, and exhibit at least one activity of a 2246 protein. Typically, biologically active portions comprise a

domain or motif with at least one activity of the 2246 protein. A biologically active portion of a 2246 protein can be a polypeptide which is, for example, at least 10, 25, 50, 100 or more amino acids in length.

In a preferred embodiment, the 2246 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 2246 protein is substantially homologous to SEQ ID NO:2, and retains the functional activity of the protein of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the 2246 protein is a protein which comprises an amino acid sequence at least about 41%, 42%, 45%, 50%, 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or more homologous to the amino acid sequence of SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (e.g., when aligning a second sequence to the 2246, amino acid sequence of SEQ ID NO:2 having 674 amino acid residues, at least about 202, preferably at least 270, more preferably at least 337, even more preferably at least 404, and even more preferably at least 472, 539 or 607 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 2246 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 2246 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides 2246 chimeric or fusion proteins. As used herein, a 2246 "chimeric protein" or "fusion protein" comprises a 2246 polypeptide operatively linked to a non-2246 polypeptide. An "2246 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to 2246, whereas a "non-2246 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 2246 protein, e.g., a protein which is different from the 2246 protein and which is derived from the same or a different organism. Within a 2246 fusion protein the 2246 polypeptide can

correspond to all or a portion of a 2246 protein. In a preferred embodiment, a 2246 fusion protein comprises at least one biologically active portion of a 2246 protein. In another preferred embodiment, a 2246 fusion protein comprises at least two biologically active portions of a 2246 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the 2246 polypeptide and the non-2246 polypeptide are fused in-frame to each other. The non-2246 polypeptide can be fused to the N-terminus or C-terminus of the 2246 polypeptide.

For example, in one embodiment, the fusion protein is a GST-2246 fusion protein in which the 2246 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 2246.

In another embodiment, the fusion protein is a 2246 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 2246 can be increased through use of a heterologous signal sequence.

The 2246 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 2246 fusion proteins can be used to affect the bioavailability of a 2246 substrate. Use of 2246 fusion proteins may be useful therapeutically for the treatment of cellular growth related disorders, e.g., cardiovascular disorders. Moreover, the 2246-fusion proteins of the invention can be used as immunogens to produce anti-2246 antibodies in a subject, to purify 2246 ligands and in screening assays to identify molecules which inhibit the interaction of 2246 with a 2246 substrate.

Preferably, a 2246 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two

consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 2246-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 2246 protein.

The present invention also pertains to variants of the 2246 proteins which function as either 2246 agonists (mimetics) or as 2246 antagonists. Variants of the 2246 proteins can be generated by mutagenesis, e.g., discrete point mutation or truncation of a 2246 protein. An agonist of the 2246 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 2246 protein. An antagonist of a 2246 protein can inhibit one or more of the activities of the naturally occurring form of the 2246 protein by, for example, competitively modulating a cardiovascular system activity of a 2246 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 2246 protein.

In one embodiment, variants of a 2246 protein which function as either 2246 agonists (mimetics) or as 2246 antagonists respectively can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 2246 protein for 2246 protein agonist or antagonist activity. In one embodiment, a variegated library of 2246 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of 2246 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential 2246 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of 2246 sequences therein. There are a variety of methods which can be used to produce libraries of potential 2246 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the

synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential 2246 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a 2246 protein coding sequence can be used to generate a variegated population of 2246 fragments respectively for screening and subsequent selection of variants of a 2246 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a 2246 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the 2246 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of 2246 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 2246 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated 2246 library. For example, a library of expression vectors can be transfected into a cell line which ordinarily synthesizes and secretes 2246. The transfected cells are then cultured such that 2246 and a particular mutant 2246 are secreted and the effect of expression of the mutant on 2246 activity in cell supernatants can be detected, e.g., by any of a number of enzymatic assays. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of 2246 activity, and the individual clones further characterized.

An isolated 2246 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind 2246 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length 2246 protein can be used or, alternatively, the invention provides antigenic peptide fragments of 2246 for use as immunogens. The antigenic peptide of 2246 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 2246 such that an antibody raised against the peptide forms a specific immune complex with 2246. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of 2246 that are located on the surface of the protein, e.g., hydrophilic regions.

A 2246 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed 2246 protein or a chemically synthesized 2246 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic 2246 preparation induces a polyclonal anti-2246 antibody response.

Accordingly, another aspect of the invention pertains to anti-2246 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that

contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as 2246. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind 2246. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of 2246. A monoclonal antibody composition thus typically displays a single binding affinity for a particular 2246 protein with which it immunoreacts.

Polyclonal anti-2246 antibodies can be prepared as described above by immunizing a suitable subject with a 2246 immunogen. The anti-2246 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized 2246. If desired, the antibody molecules directed against 2246 can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-2246 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a 2246 immunogen as described above,

and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds 2246.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-2246 monoclonal antibody (see, e.g., G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind 2246, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-2246 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with 2246 to thereby isolate immunoglobulin library members that bind 2246. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and

screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-2246 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature*

321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-2246 antibody (e.g., monoclonal antibody) can be used to isolate 2246 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-2246 antibody can facilitate the purification of natural 2246 from cells and of recombinantly produced 2246 expressed in host cells. Moreover, an anti-2246 antibody can be used to detect 2246 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the 2246 protein. Anti-2246 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a 2246 protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial

vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including

fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., 2246 proteins, mutant forms of 2246 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of 2246 proteins in prokaryotic or eukaryotic cells. For example, 2246 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in 2246 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 2246 proteins, for example. In a preferred embodiment, a 2246 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into

irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the 2246 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, 2246 proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an

RNA molecule which is antisense to 2246 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 2246 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a 2246 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) a 2246 protein. Accordingly, the invention further provides methods for producing a 2246 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a 2246 protein has been introduced) in a suitable medium such that a 2246 protein is produced. In another embodiment, the method further comprises isolating a 2246 protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which 2246-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous 2246 sequences have been introduced into their genome or homologous recombinant animals in which endogenous 2246 sequences have been altered. Such animals are useful for studying the function and/or activity of a 2246 and for identifying and/or evaluating modulators of 2246 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic

animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous 2246 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a 2246-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The 2246 cDNA sequence of SEQ ID NO:1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human 2246 gene, such as a mouse or rat 2246 gene, can be used as a transgene. Alternatively, a 2246 gene homologue, such as another 2246 family member, can be isolated based on hybridization to the 2246 cDNA sequences of SEQ ID NO:1 or SEQ ID NO:3 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a 2246 transgene to direct expression of a 2246 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a 2246 transgene in its genome and/or expression of 2246 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 2246 protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a 2246 gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the 2246 gene. The 2246 gene can be a human gene (e.g., the SEQ ID NO:1), but more preferably, is a non-human homologue of a human 2246 gene (e.g., a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1). For example, a mouse 2246 gene can be used to construct a homologous recombination vector suitable for altering an endogenous 2246 gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous 2246 gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous 2246 gene is mutated or otherwise altered but still encodes a functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous 2246 protein). In the homologous recombination vector, the altered portion of the 2246 gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the 2246 gene to allow for homologous recombination to occur between the exogenous 2246 gene carried by the vector and an endogenous 2246 gene in an embryonic stem cell. The additional flanking 2246 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced 2246 gene has homologously recombined with the endogenous 2246 gene are selected (see, e.g., Li, E. *et al.* (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of

the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Use of 2246 Molecules as Surrogate Markers

The 2246 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 2246 molecules of the invention can be detected, and can be correlated with one or more biological states in vivo. For example, the 2246 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen et al. (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The 2246 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a

biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug in vivo. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (e.g., a 2246 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-2246 antibodies can be employed in an immune-based detection system for a 2246 protein marker, or 2246-specific radiolabeled probes can be used to detect a 2246 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda et al. US 6,033,862; Hattis et al. (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The 2246 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod et al. (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 2246 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is

optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 2246 DNA can correlate with a 2246 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

V. Pharmaceutical Compositions

The 2246 nucleic acid molecules, 2246 proteins, and anti-2246 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The

parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a 2246 protein or anti-2246 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate,

polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the

dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

VI. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic). The isolated nucleic acid molecules of the invention can be used, for example, to express 2246 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect 2246 mRNA (e.g., in a biological sample) or a genetic alteration in a 2246 gene, and to modulate 2246 activity, as described further below. The 2246 proteins can be used to treat disorders characterized by insufficient or excessive production of a 2246 substrate or production

of 2246 inhibitors. In addition, the 2246 proteins can be used to screen for naturally occurring 2246 substrates, to screen for drugs or compounds which modulate 2246 activity, as well as to treat disorders characterized by insufficient or excessive production of 2246 protein or production of 2246 protein forms which have decreased or aberrant activity compared to 2246 wild type protein. Moreover, the anti-2246 antibodies of the invention can be used to detect and isolate 2246 proteins, regulate the bioavailability of 2246 proteins, and modulate 2246 activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to 2246 proteins, have a stimulatory or inhibitory effect on, for example, 2246 expression or 2246 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 2246 substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 2246 protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 2246 protein or polypeptide or biologically active portion thereof, e.g., modulate the ability of 2246 to interact with its cognate ligand. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J.*

Med. Chem. 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a 2246 target molecule (e.g., a 2246 phosphorylation substrate) with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the 2246 target molecule. Determining the ability of the test compound to modulate the activity of a 2246 target molecule can be accomplished, for example, by determining the ability of the 2246 protein to bind to or interact with the 2246 target molecule, or by determining the ability of the 2246 protein to phosphorylate the 2246 target molecule.

The ability of the 2246 protein to phosphorylate a 2246 target molecule can be determined by, for example, an *in vitro* kinase assay. Briefly, a 2246 target molecule, e.g., an immunoprecipitated 2246 target molecule from a cell line expressing such a molecule, can be incubated with the 2246 protein and radioactive ATP, e.g., [γ-^{32}P] ATP, in a buffer containing $MgCl_2$ and $MnCl_2$, e.g., 10 mM $MgCl_2$ and 5 mM $MnCl_2$. Following the incubation, the immunoprecipitated 2246 target molecule can be separated by SDS-polyacrylamide gel electrophoresis under reducing conditions, transferred to a membrane, e.g., a PVDF membrane, and autoradiographed. The appearance of detectable bands on the autoradiograph indicates that the 2246 substrate has been phosphorylated. Phosphoaminoacid analysis of the phosphorylated substrate can also be performed in order to determine which residues on the 2246 substrate are phosphorylated. Briefly, the radiophosphorylated protein band can be excised from the SDS gel and subjected to partial acid hydrolysis. The products can then be separated

by one-dimensional electrophoresis and analyzed on, for example, a phosphoimager and compared to ninhydrin-stained phosphoaminoacid standards.

Determining the ability of the 2246 protein to bind to or interact with a 2246 target molecule can be accomplished by determining direct binding. Determining the ability of the 2246 protein to bind to or interact with a 2246 target molecule can be accomplished, for example, by coupling the 2246 protein with a radioisotope or enzymatic label such that binding of the 2246 protein to a 2246 target molecule can be determined by detecting the labeled 2246 protein in a complex. For example, 2246 molecules, e.g., 2246 proteins, can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, 2246 molecules can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound to modulate the interaction between 2246 and its target molecule, without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of 2246 with its target molecule without the labeling of either 2246 or the target molecule. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between compound and receptor.

In a preferred embodiment, determining the ability of the 2246 protein to bind to or interact with a 2246 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., chloramphenicol acetyl transferase), or detecting a target-regulated cellular response.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a 2246 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 2246 protein or biologically active portion thereof is determined. Binding of the test compound to the 2246 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the 2246 protein or biologically active portion thereof with a known compound which binds 2246 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 2246 protein, wherein determining the ability of the test compound to interact with a 2246 protein comprises determining the ability of the test compound to preferentially bind to 2246 or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a 2246 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the 2246 protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a 2246 protein can be accomplished, for example, by determining the ability of the 2246 protein to bind to a 2246 target molecule by one of the methods described above for determining direct binding. Determining the ability of the 2246 protein to bind to a 2246 target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a 2246 protein can be accomplished by determining the ability of the 2246 protein to further modulate the activity of a 2246 target molecule (e.g., a 2246 mediated signal transduction pathway component). For example, the activity of

the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a 2246 protein or biologically active portion thereof with a known compound which binds the 2246 protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the 2246 protein, wherein determining the ability of the test compound to interact with the 2246 protein comprises determining the ability of the 2246 protein to preferentially bind to or modulate the activity of a 2246 target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of proteins (e.g., 2246 proteins or biologically active portions thereof, or receptors to which 2246 binds). In the case of cell-free assays in which a membrane-bound form a protein is used (e.g., a cell surface 2246 receptor) it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either 2246 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 2246 protein, or interaction of a 2246 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ 2246 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione

sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 2246 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 2246 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a 2246 protein or a 2246 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated 2246 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with 2246 protein or target molecules but which do not interfere with binding of the 2246 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or 2246 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 2246 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 2246 protein or target molecule.

In another embodiment, modulators of 2246 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of 2246 mRNA or protein in the cell is determined. The level of expression of 2246 mRNA or protein in the presence of the candidate compound is compared to the level of expression of 2246 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of 2246 expression based on this comparison. For example, when expression of 2246 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of

2246 mRNA or protein expression. Alternatively, when expression of 2246 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 2246 mRNA or protein expression. The level of 2246 mRNA or protein expression in the cells can be determined by methods described herein for detecting 2246 mRNA or protein.

In yet another aspect of the invention, the 2246 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 2246 ("2246-binding proteins" or "2246-bp") and are involved in 2246 activity. Such 2246-binding proteins are also likely to be involved in the propagation of signals by the 2246 proteins or 2246 targets as, for example, downstream elements of a 2246-mediated signaling pathway. Alternatively, such 2246-binding proteins are likely to be 2246 inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 2246 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a 2246-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 2246 protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a 2246 modulating agent, an antisense 2246 nucleic acid molecule, a 2246-specific antibody, or a 2246-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the 2246 nucleotide sequences, described herein, can be used to map the location of the 2246 genes on a chromosome. The mapping of the 2246 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, 2246 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 2246 nucleotide sequences. Computer analysis of the 2246 sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process.

These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 2246 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the 2246 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 9o, 1p, or 1v sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a

DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 2246 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The 2246 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 2246 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The 2246 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as

those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 2246 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial 2246 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the 2246 nucleotide sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1, having a length of at least 20 bases, preferably at least 30 bases.

The 2246 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain

tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 2246 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., 2246 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining 2246 protein and/or nucleic acid expression as well as 2246 activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant 2246 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with 2246 protein, nucleic acid expression or activity. For example, mutations in a 2246 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with 2246 protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of 2246 in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of 2246 protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 2246 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 2246 protein such that the presence of 2246 protein or nucleic acid is detected in the

biological sample. A preferred agent for detecting 2246 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to 2246 mRNA or genomic DNA. The nucleic acid probe can be, for example, a human 2246 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 2246 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting 2246 protein is an antibody capable of binding to 2246 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect 2246 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of 2246 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of 2246 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of 2246 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of 2246 protein include introducing into a subject a labeled anti-2246 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from

the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting 2246 protein, mRNA, or genomic DNA, such that the presence of 2246 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of 2246 protein, mRNA or genomic DNA in the control sample with the presence of 2246 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of 2246 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting 2246 protein or mRNA in a biological sample; means for determining the amount of 2246 in the sample; and means for comparing the amount of 2246 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 2246 protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant 2246 expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with 2246 protein, nucleic acid expression or activity. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant 2246 expression or activity in which a test sample is obtained from a subject and 2246 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of 2246 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant 2246 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant 2246 expression or activity. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant 2246 expression or activity in which a test sample is obtained and 2246 protein or nucleic acid expression or activity is detected (e.g., wherein the abundance of 2246 protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant 2246 expression or activity).

The methods of the invention can also be used to detect genetic alterations in a 2246 gene, thereby determining if a subject with the altered gene is at risk for a disorder associated with the 2246 gene. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 2246-protein, or the mis-expression of the 2246 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 2246 gene; 2) an addition of one or more nucleotides to a 2246 gene; 3) a substitution of one or more nucleotides of a 2246 gene, 4) a chromosomal rearrangement of a 2246 gene; 5) an alteration in the level of a messenger RNA transcript of a 2246 gene, 6) aberrant modification of a 2246 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 2246 gene, 8) a non-wild type level of a 2246 protein, 9) allelic loss of a 2246 gene, and 10) inappropriate post-translational modification of a 2246 protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a 2246 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a

ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the 2246 gene (see Abravaya *et al.* (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 2246 gene under conditions such that hybridization and amplification of the 2246 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a 2246 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in 2246 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density

arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 2246 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 2246 gene and detect mutations by comparing the sequence of the sample 2246 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the 2246 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type 2246 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with

RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 2246 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a 2246 sequence, e.g., a wild-type 2246 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 2246 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control 2246 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double

stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner *et al.* (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5'

sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 2246 gene.

Furthermore, any cell type or tissue in which 2246 is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs or compounds) on the expression or activity of a 2246 protein can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 2246 gene expression, protein levels, or upregulate 2246 activity, can be monitored in clinical trials of subjects exhibiting decreased 2246 gene expression, protein levels, or downregulated 2246 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 2246 gene expression, protein levels, or downregulate 2246 activity, can be monitored in clinical trials of subjects exhibiting increased 2246 gene expression, protein levels, or upregulated 2246 activity. In such clinical trials, the expression or activity of a 2246 gene, and preferably, other genes that have been implicated in a disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including 2246, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates 2246 activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on a 2246 associated disorder, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of 2246 and other genes implicated in the 2246 associated disorder, respectively. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as

described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of 2246 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a 2246 protein, mRNA, or genomic DNA in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the 2246 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the 2246 protein, mRNA, or genomic DNA in the pre-administration sample with the 2246 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of 2246 to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of 2246 to lower levels than detected, i.e. to decrease the effectiveness of the agent. According to such an embodiment, 2246 expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant 2246 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be

specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 2246 molecules of the present invention or 2246 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant 2246 expression or activity, by administering to the subject a 2246 or an agent which modulates 2246 expression or at least one 2246 activity. Subjects at risk for a disease which is caused or contributed to by aberrant 2246 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 2246 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 2246 aberrancy, for example, a 2246, 2246 agonist or 2246 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating 2246 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 2246 or agent that modulates one or more of the activities of 2246 protein activity associated with the cell. An agent that modulates 2246 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 2246 protein (e.g., a 2246 phosphorylation substrate), a 2246 antibody, a 2246 agonist or antagonist, a peptidomimetic of a 2246 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more 2246 activities. Examples of such stimulatory agents include active 2246 protein and a nucleic acid molecule encoding 2246 that has been introduced into the cell. In another embodiment, the agent inhibits one or more 2246 activities. Examples of such inhibitory agents include antisense 2246 nucleic acid molecules, anti-2246 antibodies, and 2246 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a 2246 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 2246 expression or activity. In another embodiment, the method involves administering a 2246 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant 2246 expression or activity.

Stimulation of 2246 activity is desirable in situations in which 2246 is abnormally downregulated and/or in which increased 2246 activity is likely to have a beneficial effect. For example, stimulation of 2246 activity is desirable in situations in which a 2246 is downregulated and/or in which increased 2246 activity is likely to have a beneficial effect. Likewise, inhibition of 2246 activity is desirable in situations in which 2246 is abnormally upregulated and/or in which decreased 2246 activity is likely to have a beneficial effect.

3. Pharmacogenomics

The 2246 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 2246 activity (e.g., 2246 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., cardiovascular disorders such as congestive heart failure) associated with aberrant 2246 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 2246 molecule or 2246 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 2246 molecule or 2246 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a

"bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict a drug response. According to this method, if a gene that encodes a drug target is known (e.g., a 2246 protein or 2246 receptor of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19

quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 2246 molecule or 2246 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 2246 molecule or 2246 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1: Identification and Characterization of Human 2246 cDNAs

The human 2246 sequence (Figure 1A-B; SEQ ID NO:1), which is approximately 2219 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence (SEQ ID NO:3) of about 2025 nucleotides. The coding sequence encodes a 674 amino acid protein (SEQ ID NO:2).

Example 2: Expression and Tissue Distribution of 2246 mRNA

Northern blot hybridizations with various RNA samples can be performed under standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65° C. A DNA probe corresponding to all or a portion of the 2246 cDNA (SEQ ID NO:1) can be used. The DNA is radioactively labeled with ^{32}P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations. TaqMan real-time quantitative RT-PCR is used to detect the presence of RNA transcript corresponding to human 2246 in several tissues. It is found that the corresponding orthologs of 2246 are expressed in a variety of tissues.

Figure 6 illustrates the relative expression levels of 2246 in various tissues using TaqMan PCR, and significant expression is found in normal human brain cortex, and high expression in pancreas and brain hypothalamus.

Reverse Transcriptase PCR (RT-PCR) is used to detect the presence of RNA transcript corresponding to human 2246 in RNA prepared from tumor and normal tissues. If a subject has a disease characterized by underexpression or overexpression of a 2246 gene, modulators which have a stimulatory or inhibitory effect on protein kinase activity (e.g., protein kinase gene expression) can be administered to individuals to treat (prophylactically or therapeutically) protein kinase-associated disorders.

2246 molecules are found to be overexpressed or underexpressed in some tumor or cells, where the molecules may be inappropriately propagating either cell proliferation or cell survival signals or have aberrant protein kinase activity. As such, 2246 molecules may serve as specific and novel identifiers of such tumor cells or disorders.

Further, modulators of the 2246 molecules are useful for the treatment of cancer. For example, inhibitors of the 2246 molecules are useful for the treatment of cancer where 2246 is upregulated in tumor cells and are useful as a diagnostic. In addition, activators of the 2246 molecules are useful for the treatment of cancer, where 2246 expression is downregulated.

Example 3: Recombinant Expression of 2246 in Bacterial Cells

In this example, 2246 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 2246 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-3714, -16742, -23546, or -13887 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 4: Expression of Recombinant 2246 Protein in COS Cells

To express the 2246 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 2246 protein and an HA tag (Wilson et al. (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 2246 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 2246 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 2246 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 2246 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La

Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 2246-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 2246 polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 2246 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 2246 polypeptide is detected by radiolabelling and immunoprecipitation using a 2246 specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. An isolated 2246 nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____;

b) a nucleic acid molecule comprising a fragment of at least 15 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or a complement thereof, under stringent conditions;

f) a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____; and

g) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____.

2. The isolated nucleic acid molecule of claim 1, which is the nucleotide sequence SEQ ID NO:1.

3. A host cell which contains the nucleic acid molecule of claim 1.

4. An isolated 2246 polypeptide selected from the group consisting of:

a) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or a complement thereof under stringent conditions;

c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2; and

d) the amino acid sequence of SEQ ID NO:2.

5. An antibody which selectively binds to a polypeptide of claim 4.

6. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the fragment

comprises at least 15 contiguous amino acids of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____;

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1 or SEQ ID NO:3; and

d) the amino acid sequence of SEQ ID NO:2;

comprising culturing the host cell of claim 3 under conditions in which the nucleic acid molecule is expressed.

7. A method for detecting the presence of a nucleic acid molecule of claim 1 or a polypeptide encoded by the nucleic acid molecule in a sample, comprising:

a) contacting the sample with a compound which selectively hybridizes to the nucleic acid molecule of claim 1 or binds to the polypeptide encoded by the nucleic acid molecule; and

b) determining whether the compound hybridizes to the nucleic acid or binds to the polypeptide in the sample.

8. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 or binds to a polypeptide encoded by the nucleic acid molecule and instructions for use.

9. A method for identifying a compound which binds to a polypeptide or modulates the activity of the polypeptide of claim 4 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 4 with a test compound; and

b) determining whether the polypeptide binds to the test compound or determining the effect of the test compound on the activity of the polypeptide.

10. A method for modulating the activity of a polypeptide of claim 4 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

11. A method of identifying a nucleic acid molecule associated with cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample from a subject with or at risk of developing cancer or a cellular proliferation and/or differentiation disorder comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 defined in claim 2; and

b) detecting the presence of a nucleic acid molecule in the sample that hybridizes to the probe, thereby identifying a nucleic acid molecule associated with cancer or a cellular proliferation and/or differentiation disorder.

12. A method of identifying a nucleic acid associated with cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample from a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk of developing a cancer or a cellular proliferation and/or differentiation disorder comprising nucleic acid molecules with a first and a second amplification primer, the first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 defined in claim 2 and the second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1;

b) incubating the sample under conditions that allow nucleic acid amplification; and

c) detecting the presence of a nucleic acid molecule in the sample that is amplified, thereby identifying the nucleic acid molecule associated with cancer or a cellular proliferation and/or differentiation disorder.

13. A method of identifying a polypeptide associated with cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample comprising polypeptides with a 2246 binding partner of the 2246 polypeptide defined in claim 4; and

b) detecting the presence of a polypeptide in the sample that binds to the 2246 binding partner, thereby identifying the polypeptide associated with cancer or a cellular proliferation and/or differentiation disorder.

14. A method of identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample obtained from the subject comprising nucleic acid molecules with a hybridization probe comprising at least 25 contiguous nucleotides of SEQ ID NO:1 defined in claim 2; and

b) detecting the presence of a nucleic acid molecule in the sample that hybridizes to the probe, thereby identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing a cancer or a cellular proliferation and/or differentiation disorder.

15. A method of identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing a cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample obtained from the subject comprising nucleic acid molecules with a first and a second amplification primer, the first primer comprising at least 25 contiguous nucleotides of SEQ ID NO:1 defined in claim 2 and the second primer comprising at least 25 contiguous nucleotides from the complement of SEQ ID NO:1;

b) incubating the sample under conditions that allow nucleic acid amplification; and

c) detecting the presence of a nucleic acid molecule in the sample that is amplified, thereby identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing cancer or a cellular proliferation and/or differentiation disorder.

16. A method of identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing cancer or a cellular proliferation and/or differentiation disorder comprising:

a) contacting a sample obtained from the subject comprising polypeptides with a 2246 binding partner of the 2246 polypeptide defined in claim 4; and

b) detecting the presence of a polypeptide in the sample that binds to the 2246 binding partner, thereby identifying a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk for developing cancer or a cellular proliferation and/or differentiation disorder.

17. A method for identifying a compound capable of treating cancer or a cellular proliferation and/or differentiation disorder characterized by aberrant 2246 nucleic acid expression or 2246 polypeptide activity comprising assaying the ability of the compound to modulate 2246 nucleic acid expression or 2246 polypeptide activity, thereby identifying a compound capable of treating cancer or a cellular proliferation and/or differentiation disorder characterized by aberrant 2246 nucleic acid expression or 2246 polypeptide activity.

18. A method for treating a subject having cancer or a cellular proliferation and/or differentiation disorder or at risk of developing cancer or a cellular proliferation and/or differentiation disorder comprising administering to the subject a 2246 modulator of the nucleic acid molecule defined in claim 1 or the polypeptide encoded by the nucleic acid molecule or contacting a cell with a 2246 modulator.

19. The method of claim 18, wherein the 2246 modulator is

a) a small molecule;

b) peptide;

c) phosphopeptide;

d) anti-2246 antibody;

e) a 2246 polypeptide comprising the amino acid sequence of SEQ ID NO:2, or a fragment thereof;

f) a 2246 polypeptide comprising an amino acid sequence which is at least 90 percent identical to the amino acid sequence of SEQ ID NO:2, wherein the percent identity is calculated using the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; or

g) an isolated naturally occurring allelic variant of a polypeptide consisting of the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 at 6X SSC at 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C.

20. The method of claim 18, wherein the 2246 modulator is

a) an antisense 2246 nucleic acid molecule;

b) is a ribozyme;

c) the nucleotide sequence of SEQ ID NO:1, or a fragment thereof;

d) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence which is at least 90 percent identical to the amino acid sequence of SEQ ID NO:2, wherein the percent identity is calculated using the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4;

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 at 6X SSC at 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; or

f) a gene therapy vector.

21. A method for evaluating the efficacy of a treatment of cancer or a cellular proliferation and/or differentiation disorder, in a subject, comprising:

treating a subject with a protocol under evaluation;

assessing the expression level of a 2246 nucleic acid molecule defined in claim 1 or 2246 polypeptide encoded by the 2246 nucleic acid molecule,

wherein a change in the expression level of 2246 nucleic acid or 2246 polypeptide after the treatment, relative to the level before the treatment, is indicative of the efficacy of the treatment of cancer or a cellular proliferation and/or differentiation disorder.

22. A method of diagnosing cancer or a cellular proliferation and/or differentiation disorder in a subject, comprising:

evaluating the expression or activity of a 2246 nucleic acid molecule defined in claim 1 or a 2246 polypeptide encoded by the 2246 nucleic acid molecule, such that a difference in the level of 2246 nucleic acid or 2246 polypeptide relative to a normal subject or a cohort of normal subjects is indicative of cancer or a cellular proliferation and/or differentiation disorder.

```
                                                                M  T  S  T  G  K        6 SEQ ID NO:2
CCACGCGTCCGGCGCCCGGGCGCCCCTGGCCGGCGCCGGGCCCCAGAGCG    ATG ACA TCG ACG GGG AAG      18 SEQ ID NO:1
                                                       ↑ SEQ ID NO:3→

 D  G  G  A  Q  H  A  Q  Y  V  G  P  Y  R  L  E  K  T  L  G    26
GAC GGC GGC GCG CAG CAC GCG CAG TAT GTT GGG CCC TAC CGG CTG GAG AAG ACG CTG GGC    78

 K  G  Q  T  G  L  V  K  L  G  V  H  C  V  T  C  Q  K  V  A    46
AAG GGG CAG ACA GGT CTG GTG AAG CTG GGG GTT CAC TGC GTC ACC TGC CAG AAG GTG GCC   138

 I  K  I  V  N  R  E  K  L  S  E  S  V  L  M  K  V  E  R  E    66
ATC AAG ATC GTC AAC CGT GAG AAG CTC AGC GAG TCG GTG CTG ATG AAG GTG GAG CGG GAG   198

 I  A  I  L  K  L  I  E  H  P  H  V  L  K  L  H  D  V  Y  E    86
ATC GCG ATC CTG AAG CTC ATT GAG CAC CCC CAC GTC CTA AAG CTG CAC GAC GTT TAT GAA   258

 N  K  K  Y  L  Y  L  V  L  E  H  V  S  G  G  E  L  F  D  Y   106
AAC AAA AAA TAT TTG TAC CTG GTG CTA GAA CAC GTG TCA GGT GGT GAG CTC TTC GAC TAC   318

 L  V  K  K  G  R  L  T  P  K  E  A  R  K  F  F  R  Q  I  I   126
CTG GTG AAG AAG GGG AGG CTG ACG CCT AAG GAG GCT CGG AAG TTC TTC CGG CAG ATC ATC   378

 S  A  L  D  F  C  H  S  H  S  I  C  H  R  D  L  K  P  E  N   146
TCT GCG CTG GAC TTC TGC CAC AGC CAC TCC ATA TGC CAC AGG GAT CTG AAA CCT GAA AAC   438

 L  L  L  D  E  K  N  N  I  R  I  A  D  F  G  M  A  S  L  Q   166
CTC CTG CTG GAC GAG AAG AAC AAC ATC CGC ATC GCA GAC TTT GGC ATG GCG TCC CTG CAG   498

 V  G  D  S  L  L  E  T  S  C  G  S  P  H  Y  A  C  P  E  V   186
GTT GGC GAC AGC CTG TTG GAG ACC AGC TGT GGG TCC CCC CAC TAC GCC TGC CCC GAG GTG   558

 I  R  G  E  K  Y  D  G  R  K  A  D  V  W  S  C  G  V  I  L   206
ATC CGG GGG GAG AAG TAT GAC GGC CGG AAG GCG GAC GTG TGG AGC TGC GGC GTC ATC CTG   618

 F  A  L  L  V  G  A  L  P  F  D  D  D  N  L  R  Q  L  L  E   226
TTC GCC TTG CTG GTG GGG GCT CTG CCC TTC GAC GAT GAC AAC TTG CGA CAG CTG CTG GAG   678

 K  V  K  R  G  V  F  H  M  P  H  F  I  P  P  D  C  Q  S  L   246
AAG GTG AAG CGG GGC GTG TTC CAC ATG CCG CAC TTT ATC CCG CCC GAC TGC CAG AGT CTG   738

 L  R  G  M  S  E  V  D  A  A  R  R  L  T  L  E  H  I  Q  K   266
CTA CGG GGC ATG AGC GAG GTG GAC GCC GCA CGC CGC CTC ACG CTA GAG CAC ATT CAG AAA   798

 H  I  W  Y  I  G  G  K  N  E  P  E  P  E  Q  P  I  P  R  K   286
CAC ATA TGG TAT ATA GGG GGC AAG AAT GAG CCC GAA CCA GAG CAG CCC ATT CCT CGC AAG   858

 V  Q  I  R  S  L  P  S  L  E  D  I  D  P  D  V  L  D  S  M   306
GTG CAG ATC CGC TCG CTG CCC AGC CTG GAG GAC ATC GAC CCC GAC GTG CTG GAC AGC ATG   918

 H  S  L  G  C  F  R  D  R  N  K  L  L  Q  D  L  L  S  E  E   326
CAC TCA CTG GGC TGC TTC CGA GAC CGC AAC AAG CTG CTG CAG GAC CTG CTG TCC GAG GAG   978

 E  N  Q  E  K  M  I  Y  F  L  L  L  D  R  K  E  R  Y  P  S   346
GAG AAC CAG GAG AAG ATG ATT TAC TTC CTC CTC CTG GAC CGG AAA GAA AGG TAC CCG AGC  1038
```

FIG. 1A

```
 Q  E  D  E  D  L  P  P  R  N  E  I  D  P  P  R  K  R  V  D   366
CAG GAG GAT GAG GAC CTG CCC CCC CGG AAC GAG ATA GAC CCT CCC CGG AAG CGT GTG GAC  1098

 S  P  M  L  N  R  H  G  K  R  R  P  E  R  K  S  M  E  V  L   386
TCC CCG ATG CTG AAC CGG CAC GGC AAG CGG CGG CCA GAA CGC AAA TCC ATG GAG GTG CTC  1158

 S  V  T  D  G  G  S  P  V  P  A  R  R  A  I  E  M  A  Q  H   406
AGC GTG ACG GAC GGC GGC TCC CCG GTG CCT GCG CGG CGG GCC ATT GAG ATG GCC CAG CAC  1218

 G  Q  R  S  R  S  I  S  G  A  S  S  G  L  S  T  S  P  L  S   426
GGC CAG AGG TCT CGG TCC ATC AGC GGT GCC TCC TCA GGC CTT TCC ACC AGC CCA CTC AGC  1278

 S  P  R  V  T  P  H  P  S  P  R  G  S  P  L  P  T  P  K  G   446
AGC CCC CGG GTG ACC CCT CAC CCC TCA CCA AGG GGC AGT CCC CTC CCC ACC CCC AAG GGG  1338

 T  P  V  H  T  P  K  E  S  P  A  G  T  P  N  P  T  P  P  S   466
ACA CCT GTC CAC ACG CCA AAG GAG AGC CCG GCT GGC ACG CCC AAC CCC ACG CCC CCG TCC  1398

 S  P  S  V  G  G  V  P  W  R  A  R  L  N  S  I  K  N  S  F   486
AGC CCC AGC GTC GGA GGG GTG CCC TGG AGG GCG CGG CTC AAC TCC ATC AAG AAC AGC TTT  1458

 L  G  S  P  R  F  H  R  R  K  L  Q  V  P  T  P  E  E  M  S   506
CTG GGC TCA CCC CGC TTC CAC CGC CGG AAA CTG CAA GTT CCG ACG CCG GAG GAG ATG TCC  1518

 N  L  T  P  E  S  S  P  E  L  A  K  K  S  W  F  G  N  F  I   526
AAC CTG ACA CCA GAG TCG TCC CCA GAG CTG GCG AAG AAG TCC TGG TTT GGG AAC TTC ATC  1578

 S  L  E  K  E  E  Q  I  F  V  V  I  K  D  K  P  L  S  S  I   546
AGC CTG GAG AAG GAG GAG CAG ATC TTC GTG GTC ATC AAA GAC AAA CCT CTG AGC TCC ATC  1638

 K  A  D  I  V  H  A  F  L  S  I  P  S  L  S  H  S  V  I  S   566
AAG GCT GAC ATC GTG CAC GCC TTC CTG TCG ATT CCC AGT CTC AGC CAC AGC GTC ATC TCC  1698

 Q  T  S  F  R  A  E  Y  K  A  T  G  G  P  A  V  F  Q  K  P   586
CAA ACG AGC TTC CGG GCC GAG TAC AAG GCC ACG GGG GGG CCA GCC GTG TTC CAG AAG CCG  1758

 V  K  F  Q  V  D  I  T  Y  T  E  G  G  E  A  Q  K  E  N  G   606
GTC AAG TTC CAG GTT GAT ATC ACC TAC ACG GAG GGT GGG GAG GCG CAG AAG GAG AAC GGC  1818

 I  Y  S  V  T  F  T  L  L  S  G  P  S  R  R  F  K  R  V  V   626
ATC TAC TCC GTC ACC TTC ACC CTG CTC TCA GGC CCC AGC CGT CGC TTC AAG AGG GTG GTG  1878

 E  T  I  Q  A  Q  L  L  S  T  H  D  P  P  A  A  Q  H  L  S   646
GAG ACC ATC CAG GCC CAG CTG CTG AGC ACA CAC GAC CCG CCT GCG GCC CAG CAC TTG TCA  1938

 E  P  P  P  P  A  P  G  L  S  W  G  A  G  L  K  G  Q  K  V   666
GAA CCC CCC CCA CCA GCG CCA GGA CTA AGC TGG GGT GCT GGG CTT AAG GGC CAG AAG GTG  1998

 A  T  S  Y  E  S  S  L  *                                    675
GCC ACC AGC TAC GAG AGT AGC CTC TGA                           2025
                ← SEQ ID NO:3 →
CGCTGGCAGACACCACTAACTGTATGGAAATGATGACGGGGCGGCTTTCCAAATGTGACGAGAAGAACGGGCAGGCGGC

CCAGGCCCCCAGCACGCCCGCCAAGCGGAGTGCCCACGGCCCACTCGGTGACTCCGCGGCCGC
```

FIG. 1B

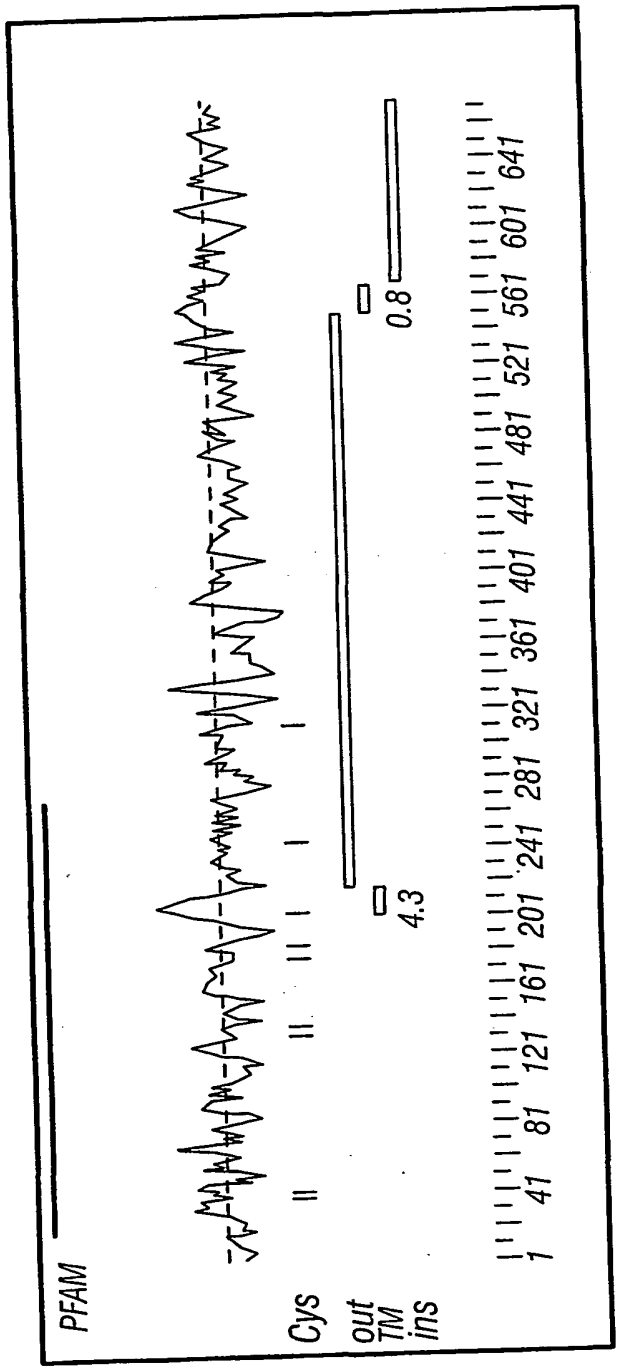


FIG. 2

```
            *->yelleklGeGsfGkVykakhk.tgkivAvKilkkesls.......lr   SEQ ID NO:4
               y+l ++lG+G  G V++++h  t ++vA+Ki+++e+ls++   +  r
2246     19    YRLEKTLGKGQTGLVKLGVHCvTCQKVAIKIVNREKLSesvlmkvER 65

               EiqilkrlsHpNIvrllgvfedtddhlylvmEymegGdLfdylrrngpls
               Ei+ilk + Hp++++l++v+e  +++lylv+E++ gG+Lfdyl+++g+l+
2246     66 EIAILKLIEHPHVLKLHDVYE-NKKYLYLVLEHVSGGELFDYLVKKGRLT 114

               ekeakkialQilrGleYLHsngivHRDLKpeNILldengtvKiaDFGLAr
               +kea+k+++Qi ++l+++Hs++i+HRDLKpeN+Llde+++++iaDFG+A
2246    115 PKEARKFFRQIISALDFCHSHSICHRDLKPENLLLDEKNNIRIADFGMAS 164

               ll...eklttfvGTpwYmmAPEvileg.rgysskvDvWSlGviLyElltg
                l  +++ l t +G+p+Y  PEv ++g++++++k+DvWS+Gvil+ ll g
2246    165 LQvgdSLLETSCGSPHYA-CPEV-IRGeKYDGRKADVWSCGVILFALLVG 212

               gplfpgadlpaftggdevdqliifvlklPfsdelpktridpleelfrikk
                                          lpf+d       d+l++l++ +k
2246    213 -------------------------ALPFDD-------DNLRQLLEKVK 229

               r.rlplpsncSeelkdLlkkcLnkDPskRpGsatakeilnhpwf<-*
               r+ +  p+ +++++++Ll+++ ++D+++R+   t ++i +h+w
2246    230 RgVFHMPHFIPPDCQSLLRGMSEVDAARRL---TLEHIQKHIWY    270
```

FIG. 3

```
Query:   282 PIPRKVQIRSLPSLEDIDPDVLDSMHSLGCFRDRNKLLQDLLSEEENQEKMIYFLLLDRK 341
             P+  +   +P  + IDPDVL  M+ LGCF ++ KL+ +L++ E N EKM+Y++LLDRK

Sbjct:     1 PMSPVIDCHPIPGEDSIDPDVLRHMNCLGCFCNKQKLINNLITPEHNTEKMVYYMLLDRK 60   SEQ ID NO:5

Query:   342 ERYPSQEDE-DL---PPRNEIDPPRKRVDSPMLNRH 373
              RYP+ +D+ D+    P    DPPRKR DS    R+

Sbjct:    61 RRYPAFDDDTDILLRGPCQHPDPPRKRTDSTRTLRY 96
```

FIG. 4

```
Query:   475 WRARLNSIKNSFLGSPRFHRRKLQVPTPEEMSN 507
             W  R     +N F  SPR HRRK Q P  ++  N
Sbjct:   195 WHQRDAQERNEFRDSPRGHRRKEQKPRSDDSEN 227   SEQ ID NO:6
```

FIG. 4B

```
Query:  521 WFGNFIS-----LEKEEQI--FVVIKDKPLSSIKADIVHAFLSIPSLSHSVISQTSFRAE 573
            WFGNF+S      E  +++   +  +++ L+SIKA+++HAFL I +L+HS++    FR +
Sbjct:    1 WFGNFMSSRYSSTEHCDELPHCIPYQNRTLNSIKAELIHAFLQIHNLTHSMVGPNCFRCD 60  SEQ ID NO:7

Query:  574 YKA--TGGPAVF-QKPVKFQVDIT------YTEGGEAQKEN--GIYSVTFTLLSGPSRRF 622
            Y+   T G +VF Q+ +KF VDI        E GE       G Y++ F+L++GP RR+
Sbjct:   61 YRRGPTSGGSVFHQRGIKFNVDIIPHSPQDRQENGEKPTSQTVGSYTIQFSLIAGPIRRY 120

Query:  623 KRVVE 627
            KR++E
Sbjct:  121 KRLLE 125
```

FIG. 5

```
Query:  113 LTPKEARKFFRQIISALDFCHS----HS-ICHR----DLKPENLLLDEK------NNIR- 156
            L+  +   +  QI   L++ HS    H  I HR    DLKPEN+LLDE+      N I+
Sbjct:  183 LSHSQLMHYVHQIAKGLEYLHSKNQKHQGIIHRAKKVDLKPENILLDEESHENTPNMIKL 242  SEQ ID NO:8

Query:  157 IADFGMA 163
            IADFG+A
Sbjct:  243 IADFGLA 249
```

FIG. 6A

```
Query:  178 SPHYACPE--VIRGEKYDGRKADVWSCGVILFALLVGALPF 216
            +P Y  PE       +KY   K+DVWS GVIL+ +L G  PF
Sbjct:  291 APEYMAPESSATNYQKYS-TKSDVWSFGVILYEMLTGKPPF 330 SEQ ID NO:9
```

FIG. 6B

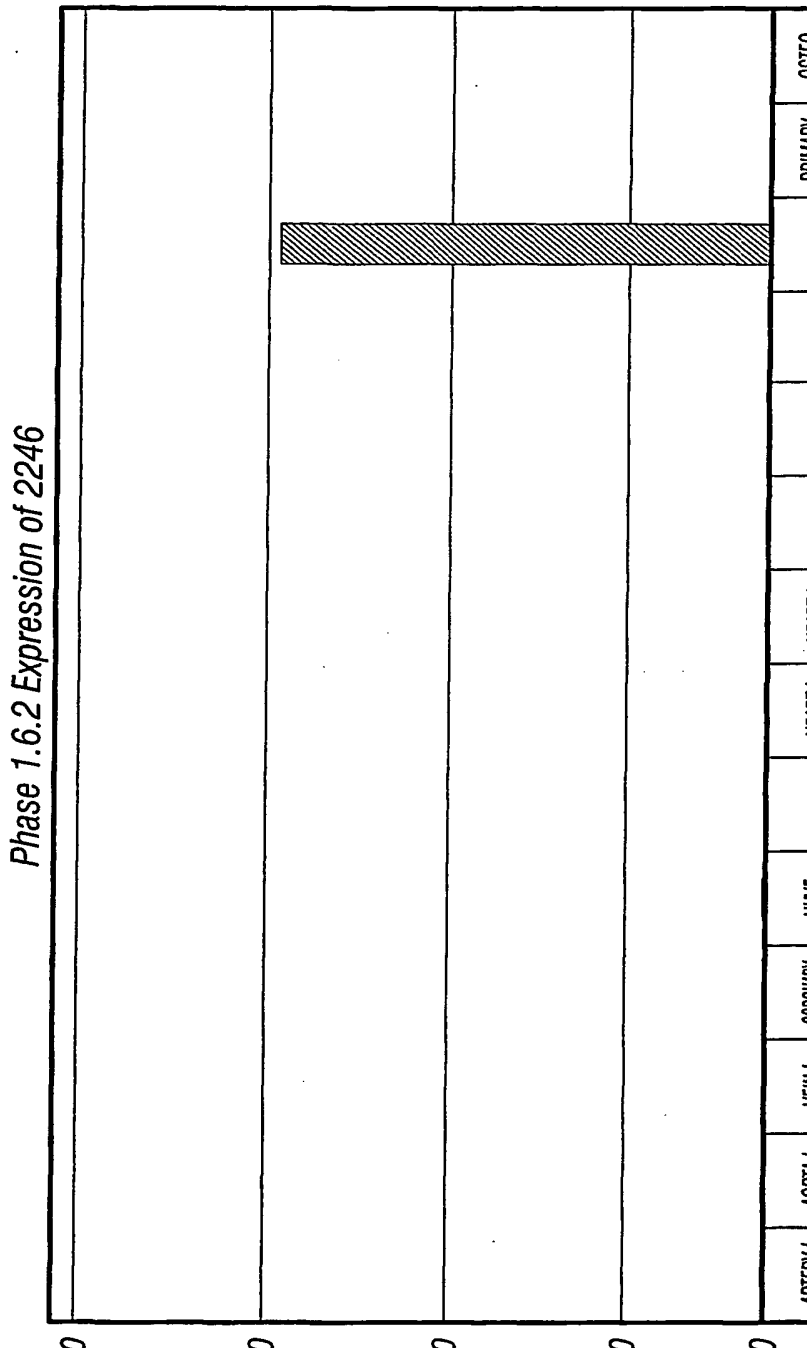
```
Query:  21 LEKTLGKGQTGLVKLGVH---CVTCQKVAIKIVNREKL-----SESVLMKVER-----EI 67
           L+K LGKG  G V    H     T + VA+K++ ++K+      ES   K E      E+
Sbjct:  37 LKKLLGKGSFGKVYKAKHKSTSTTGEVVAVKVMKKKKVMEKSSKESSSKKKEEFLMKEEV 96 SEQ ID NO.10

Query:  68 AILKLIEHPHVLKLHDVYEN-KKYLYLVLEHV 98
           + KL +H +++KL  V       +YLV E++
Sbjct:  97 VMKKLQKHVNIVKLLGVCTPFHPNIYLVYEYM 128
```

FIG. 6C

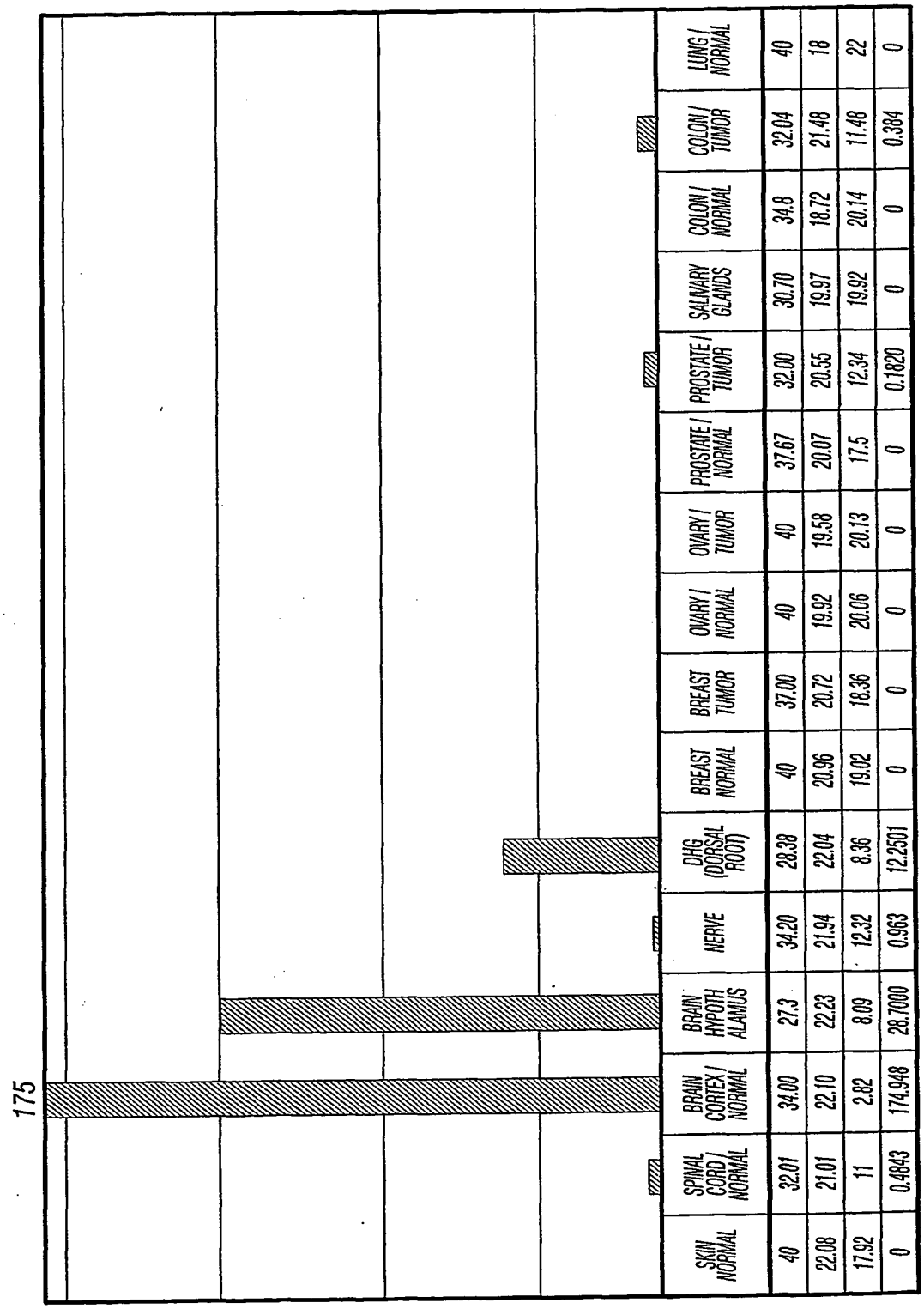
```
Query:  239 IPPDCQSLLRGMSEVDAA-RRLTLEHIQKHIWYI 271
            +  + + LL+   + D   RR T E I +H W++
Sbjct:  386 VSQEAKDLLKKCLQKDPEKRRPTFEEILQHPWFL 419  SEQ ID NO:11
```

FIG. 6D



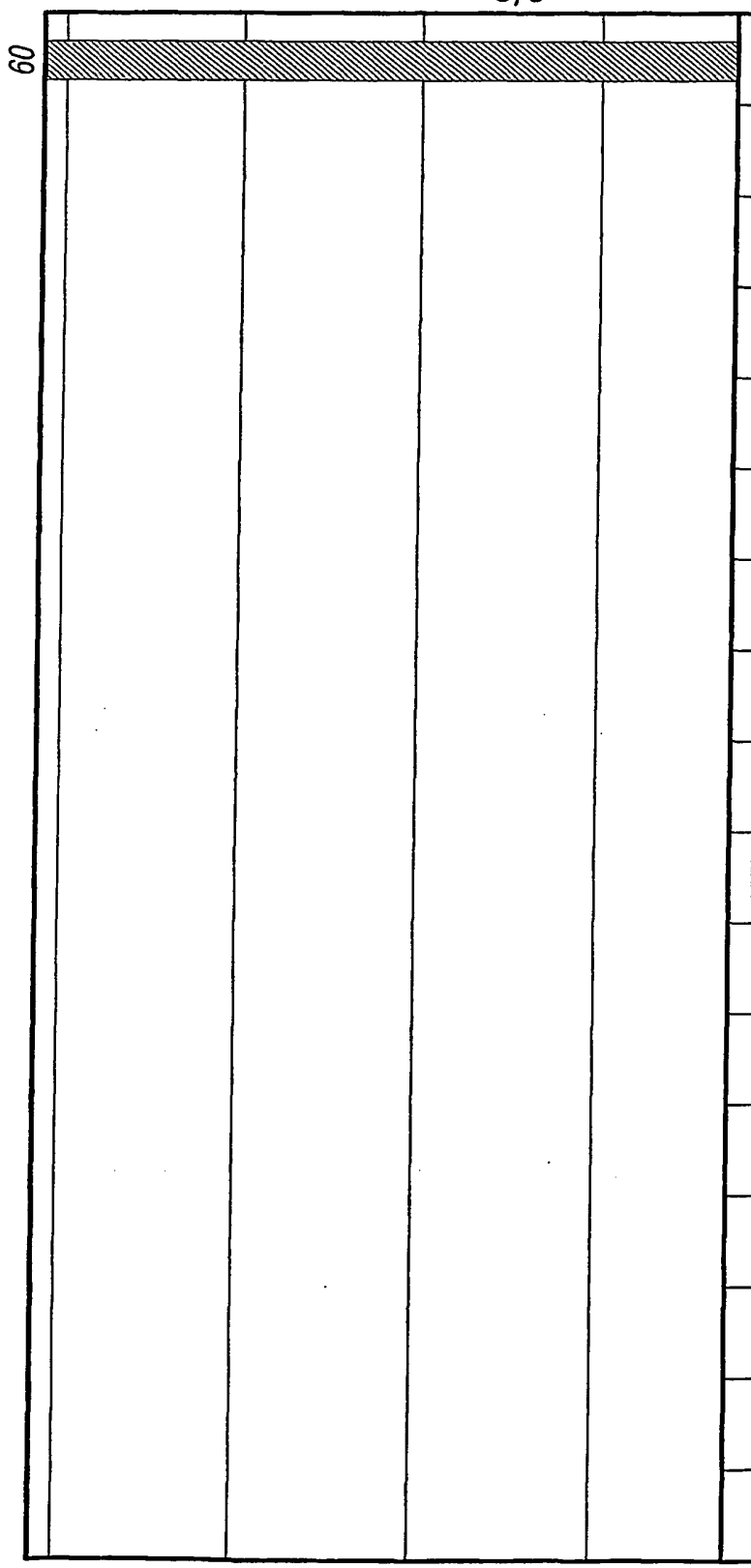
	ARTERY / NORMAL	AORTA / DISEASED	VEIN / NORMAL	CORONARY SMC	HUVE C	HEMANGIOMA	HEART / NORMAL	HEART / CHF	KIDNEY	SKELETAL MUSCLE	ADIPOSE / NORMAL	PANCREAS	PRIMARY OSTEO BLASTS	OSTEO CLASTS (DIFF)
MEAN1	40	40	40	40	34.44	40	38.86	39.35	40	40	40	27.2	40	34.03
2 MEAN	22.22	22.43	20.23	23.43	21.17	21.16	20.52	19.57	20.16	28.04	20.75	22.11	20.64	17.63
88 CT	17.78	17.87	19.77	16.57	16.27	10.02	18.34	19.79	19.95	14.96	19.25	8.06	19.16	19.41
EXPRESSION	0	0	0	0	0	0	0	0	0	0	0	29.5643	0	0

FIG. 7A



SKIN NORMAL	SPINAL CORD / NORMAL	BRAIN CORTEX / NORMAL	BRAIN HYPOTH ALAMUS	NERVE	DHG (DORSAL ROOT)	BREAST NORMAL	BREAST TUMOR	OVARY / NORMAL	OVARY / TUMOR	PROSTATE / NORMAL	PROSTATE / TUMOR	SALIVARY GLANDS	COLON / NORMAL	COLON / TUMOR	LUNG / NORMAL
40	32.01	34.00	27.3	34.20	28.38	40	37.00	40	40	37.67	32.00	30.70	34.8	32.04	40
22.08	21.01	22.10	22.23	21.94	22.04	20.96	20.72	19.92	19.58	20.07	20.55	19.97	18.72	21.48	18
17.92	11	2.82	8.09	12.32	8.36	19.02	18.36	20.06	20.13	17.5	12.34	19.92	20.14	11.48	22
0	0.4843	174.948	28.7000	0.963	12.2501	0	0	0	0	0	0.1820	0	0	0.384	0

FIG. 7B



LUNG / TUMOR	LUNG / COPD	COLON / IBD	LIVER / NORMAL	LIVER FIBROSIS	SPLEEN / NORMAL	TONSIL / NORMAL	LYMPH NODE NORMAL	SMALL INTESTINE NORMAL	MACROPHAGE	SYNOVIUM	BM-MHC	ACTIVATED PBMC	NEUTRO PHILS	MEGA KARYOCYTES	ERYTHROID	POSITIVE CONTROL
34.12	40	39.96	40	40	40	40	30.24	34.38	40	40	40	30.24	40	40	40	24.26
20.32	18.48	19.36	20.15	20.83	19.9	17.23	19.6	20.32	17.14	19.6	19.05	18.06	18.9	18.73	21.96	20.23
16.2	21.52	22.42	19.85	19.17	20.1	22.77	19.45	18.07	22.96	20.2	21.14	21.04	21.1	21.27	18.36	4.06
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60.371

FIG. 7C

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 November 2001 (01.11.2001)

PCT

(10) International Publication Number
WO 01/81588 A3

(51) International Patent Classification[7]: C12N 15/54, 9/12, 1/21, C07K 16/40, C12Q 1/68, G01N 33/68, A61K 38/43, 39/395, 31/7088

(21) International Application Number: PCT/US01/13784

(22) International Filing Date: 25 April 2001 (25.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/199,391 25 April 2000 (25.04.2000) US

(71) Applicant *(for all designated States except US)*: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 01239 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: MEYERS, Rachel [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

(74) Agents: FAVORITO, Carolyn, A. et al.; Morrison & Foerster LLP, Suite 500, 3811 Valley Centre Drive, San Diego, CA 92130-2332 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: 2246, NOVEL PROTEIN KINASE MOLECULES AND USES THEREFOR

```
          *->yelleklGeGsfGkVykakhk.tgkivAvKilkkesls.......lr  SEQ ID NO:4
             y+l ++lG+G  G V++++h  t ++vA+Ki+++e+ls++   + r
2246   19    YRLEKTLGKGQTGLVKLGVHCvTCQKVAIKIVNREKLSesvlmkvER 65

          EiqilkrlsdpnivrllgvfedtddhlylvmEymegGdLfdylrrngpls
          Ei+ilk + Hp++++l++v+e  +++lylv+E++ gG+Lfdyl+++g+l+
2246   66 EIAILKLIEHPHVLKLHDVYE-NKKYLYLVLEHVSGGELFDYLVKKGRLT 114

          ekeakkialQilrGleYLHsngivHRDLKpeNILldengtvKiaDFGLAr
          +kea+k+++Qi ++l+++Hs++i+HRDLKpeN+Llde+++++iaDFG+A
2246   115 PKEARKFFRQIISALDFCHSHSICHRDLKPENLLLDEKNNIRIADFGMAS 164

          ll...eklttfvGTpwYmmAPEvileg.rgysskvDvWSlGviLyElltg
           l  +++ l t +G+p+Y  PEv ++g++++++k+DvWS+Gvil+ ll g
2246   165 LQvgdSLLETSCGSPHYA-CPEV-IRGeKYDGRKADVWSCGVILFALLVG 212

          gplfpgadlpaftggdevdqliifvlklPfsdelpktridpleelfrikk
                                      lpf+d       d+l++l++ +k
2246   213 ----------------------------ALPFDD-------DNLRQLLEKVK 229

          r.rlplpsncSeelkdLlkkcLnkDPskRpGsatakeilnhpwf<-*
          r+ +  p+ +++++++Ll+++ ++D+++R+   t ++i +h+w
2246   230 RgVFHMPHFIPPDCQSLLRGMSEVDAARRL---TLEHIQKHIWY    270
```

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 2246 nucleic acid molecules, which encode novel protein kinases. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 2246 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which 2246 gene has been introduced or disrupted. The invention still further provides isolated 2246 proteins, fusion proteins, antigenic peptides and anti-2246 antibodies. Diagnostic, screening and therapeutic methods utilizintg compositions of the invention are also provided.

WO 01/81588 A3

(88) Date of publication of the international search report:
4 April 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 01/13784

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/54 C12N9/12 C12N1/21 C07K16/40 C12Q1/68
G01N33/68 A61K38/43 A61K39/395 A61K31/7088

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE E;BL SEQUENCE DATABSE 'Online! Hinxton, UK; 8 June 1998 (1998-06-08) STANCHI F: "Homo sapiens mRNA for putative serine/threonine kinase, partial" XP002186054 EMBL Sequence Accession no. AJ006701; abstract & YEAST, vol. 18, 2001, pages 69-80, WILEY, CHICHESTER, UK --- -/--	1-9

[X] Further documents are listed in the continuation of box C

[X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 December 2001	16/01/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HANKS S K ET AL: "THE EUKARYOTIC PROTEIN KINASE SUPERFAMILY: KINASE (CATALYTIC) DOMAIN STRUCTURE AND CLASSIFICATION" FASEB JOURNAL, FED. OF AMERICAN SOC. FOR EXPERIMENTAL BIOLOGY, BETHESDA, MD, US, vol. 9, 1 May 1995 (1995-05-01), pages 576-596, XP002062374 ISSN: 0892-6638 the whole document	
A	WO 00 15770 A (INCYTE PHARMA INC ;CORLEY NEIL C (US); BANDMAN OLGA (US); GOLI SUR) 23 March 2000 (2000-03-23) the whole document	
A	US 5 985 635 A (BANDMAN OLGA ET AL) 16 November 1999 (1999-11-16) the whole document	
A	WO 98 11234 A (HAWKINS PHILLIP R ;INCYTE PHARMA INC (US); AU YOUNG JANICE (US); G) 19 March 1998 (1998-03-19) the whole document	
A	WO 00 14212 A (ACTON SUSAN ;MILLENNIUM PHARM INC (US)) 16 March 2000 (2000-03-16) the whole document	
A	SEN S ET AL: "A PUTATIVE SERINE/THREONINE KINASE ENCODING GENE BTAK ON CHROMOSOME20Q13 IS AMPLIFIED AND OVEREXPRESSED IN HUMAN BREAST CANCER CELL LINES" ONCOGENE, BASINGSTOKE, HANTS, GB, vol. 14, 1997, pages 2195-2200, XP002917804 ISSN: 0950-9232 the whole document	
A	DONG S M ET AL: "FREQUENT SOMATIC MUTATIONS IN SERINE/THREONINE KINASE 11/PEUTZ-JEGHERS SYNDROME GENE IN LEFT-SIDED COLON CANCER" CANCER RESEARCH, AMERICAN ASSOCIATION FOR CANCER RESEARCH, BALTIMORE, MD, US, vol. 58, no. 17, 1 September 1998 (1998-09-01), pages 3787-3790, XP000938616 ISSN: 0008-5472 the whole document	

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHENG J Q ET AL: "AKT2, A PUTATIVE ONCOGENE ENCODING A MEMBER OF A SUBFAMILY OF PROTEIN-SERINA/THREONINE KINASES, IS AMPLIFIED IN HUMAN OVARIAN CARCINOMAS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 89, 1989, pages 9267-9271, XP000917044 ISSN: 0027-8424 the whole document	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims (10,18-20)-partially (as far as in vivo methods are concerned) are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 21 and 22 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: (10,18,19)-partially

Present claim 10,18 and 19 relate to a method of modulating the activity of the 2246-polypeptide or an 2246-expressing cell, respectively a method of treating a subject having cancer by using an agent that modulates the activity or expression of said 2246-polypeptide without giving a true technical characterization. Moreover no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.
Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to antibodies, antisense molecules, ribozymes, polypeptides and nucleic acids, the structure of which can be directly derived from SEQ ID Nos. 1-3.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte ional Application No
PC /US 01/13784

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0015770	A	23-03-2000	US	6156523 A	05-12-2000
			AU	6049399 A	03-04-2000
			EP	1114146 A2	11-07-2001
			WO	0015770 A2	23-03-2000
US 5985635	A	16-11-1999	US	6156523 A	05-12-2000
WO 9811234	A	19-03-1998	US	5773699 A	30-06-1998
			US	5863780 A	26-01-1999
			WO	9811234 A2	19-03-1998
			AU	4261197 A	02-04-1998
			EP	0927257 A2	07-07-1999
			JP	2001506843 T	29-05-2001
			US	6045792 A	04-04-2000
			US	6232077 B1	15-05-2001
WO 0014212	A	16-03-2000	US	6183962 B1	06-02-2001
			AU	5817799 A	27-03-2000
			EP	1112354 A1	04-07-2001
			WO	0014212 A1	16-03-2000
			US	6043040 A	28-03-2000
			US	6146841 A	14-11-2000
			US	6180358 B1	30-01-2001
			US	6153417 A	28-11-2000
			US	6146832 A	14-11-2000
			US	6190874 B1	20-02-2001
			US	6121030 A	19-09-2000
			US	6200770 B1	13-03-2001
			US	6214597 B1	10-04-2001